FAR-INFRARED LASER INVESTIGATIONS OF THE HIGH-FREQUENCY CONDUCTIVITY OF METALS UNDER QUANTUM CONDITIONS: SEMIMETALS IN HIGH MAGNETIC FIELDS AND POINT CONTACTS

R.W. VAN DER HEIJDEN

FAR-INFRARED LASER INVESTIGATIONS OF THE HIGH-FREQUENCY CONDUCTIVITY OF METALS UNDER QUANTUM CONDITIONS: SEMIMETALS IN HIGH MAGNETIC FIELDS AND POINT CONTACTS

PROMOTOR:
PROF. DR. P. WYDER

FAR-INFRARED LASER INVESTIGATIONS OF THE HIGH-FREQUENCY CONDUCTIVITY OF METALS UNDER QUANTUM CONDITIONS: SEMIMETALS IN HIGH MAGNETIC FIELDS AND POINT CONTACTS

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE
WISKUNDE EN NATUURWETENSCHAPPEN
AAN DE KATHOLIEKE UNIVERSITEIT TE NIJMEGEN, OP GEZAG VAN
DE RECTOR MAGNIFICUS PROF. DR. J.H.G.I. GIESBERS,
VOLGENS BESLUIT VAN HET COLLEGE VAN DEKANEN
IN HET OPENBAAR TE VERDEDIGEN
OP VRIJDAG 8 OKTOBER 1982
DES NAMIDDAGS TE 4.00 UUR

door

ROBERTUS WILLEBRORDUS VAN DER HEIJDEN
geboren te Oss

1982
Druk: Krips Repro Meppel

Bij het tot stand komen van dit proefschrift heb ik bijzonder veel te danken gehad aan Hans Stoelinga die mij, tot aan zijn overlijden, met inspirerende en stimulerende diskussies en met zijn deskundige adviezen, bijzonder heeft geholpen.

Verder dank ik alle (oud-)medewerkers en studenten van de afdeling Experimentele Natuurkunde 4, voor de prettige sfeer en goede samenwerking.

Voorts bedank ik alle dienstverlenende afdelingen van de Fakulteit der Wiskunde en Natuurwetenschappen voor hun deskundige en steeds bereidwillige medewerking, met name: Instrument- en Service Instrumentmakerij, Glasinstrumentmakerij, Elektronika, Ontwerp, Reprografie, Illustratie en Fotografie.

Graag zou ik de volgende personen nog met name willen noemen: Harrie Swartjes, voor de uitstekende samenwerking en zijn belangrijke bijdrage aan hoofdstuk IV en V; Hans Bluyssen en Albert van Etteger, voor hun hulp en medewerking bij de ver-infrarood laser; Louis Jansen, voor zijn introduktie in de puntkontakttechniek en de waardevolle diskussies; Ap van Gelder, voor de talloze leerzame diskussies; Herman van Kempen, voor zijn kritische beschouwing van het manuscript; Kees Beers en Jan Gerritsen, voor velerlei hulp en technische bijstand, evenals Jan Hermsen, die ik bovendien bedank voor de ingenieuze ontwerpen; Louis Schreurs, voor het beschikbaarstellen en bewerken van de Bi- en Sb samples; Martien van de Steeg, voor de computerhulp en het gebruik van Plot 81 e.d.; de vroegere en tegenwoordige "laser"-kollega's, Jan Kees Maan, Hans Sigg en Jos Verschueren voor de prettige samenwerking; Riki Gommers, voor velerlei hulp op administratief gebied; Lies Dreuning, voor de medeverzorging van het manuscript.

Dit onderzoek werd uitgevoerd op de afdeling Experimentele Natuurkunde 4 van het Research Instituut voor Materialen van de Katholieke Universiteit te Nijmegen, onder leiding van Prof.Dr. P. Wyder. Een gedeelte van dit onderzoek is gesteund door de Stichting voor Fundamenteel Onderzoek der Materie (FOM) met financiële bijdragen van de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (ZWO).

We acknowledge the permission to reprint previously published papers, obtained from the publishers of Applied Physics Letters and Solid State Communications.

CONTENTS

CHAPTER I. GENERAL INTRODUCTION.

Because of its basic and fundamental importance, together with the profound impact on technology, the study of the electrical conductivity of solids continues to be a major topic in solid state physics over more than a century. Current and actual efforts range from the study of "simple" electrical transport in "simple" prototype metals under extreme conditions (preparation, accuracy, temperature, magnetic fields, etc.), to the investigation of electrical properties of newmade materials and devices (special alloys, one- and two-dimensional systems, junctions, superlattices, organic metals, amorphous metals, etc.).

Despite the strong interaction of electrons with impurities and the lattice, the free electron theory of metals is known to hold remarkably well within the relaxation time approximation (Drude theory), if the effective electron mass is allowed to be different from the free electron mass. In this concept, the interaction of the electrons with the lattice, including imperfections, is described in terms of a mean free lifetime τ for the electrons only. An important tool in the study of metals and semiconductors is the investigation of the ac conductivity at frequencies larger than the relaxation frequency, or $\omega\tau > 1$ (preferably in combination with dc magnetic fields). Under these conditions, the ac conductivity differs essentially from the dc conductivity. The condition $\omega\tau > 1$ is satisfied at far-infrared (FIR) frequencies ($10^2 - 10^4$ GHz) for a relatively large class of materials and circumstances. For this reason, the FIR-range is well established in the spectroscopy of electronic systems in metals, semimetals and semiconductors. The FIR-range is moreover particularly relevant because many important excitation energies are found in this range. The typical phonon energies, which determine the electrical conductivity in ordinary metals, the magnetic energy level separation of a strongly quantized electronic system (Landau levels), as well as the superconducting energy gap of a superconductor are situated in the FIR. For selected materials (e.g. the semimetal Bi), the Fermi-energy and the plasma frequency are found in the FIR as well. All these conditions characterize to a great extent the experiments described in this work.

In this thesis, experimental results are reported of the far-infrared electrical conductivity of two simple physical systems, which have been selected because of their common property that they are able to display quantum effects in the conductivity under proper conditions. The term quantum effects is used here to denote those effects which essentially go beyond the Drude-like theories. The first system consists of a bulk piece of semimetal, whose electronic energy level scheme is strongly quantized by the application of a strong magnetic field. High frequency (hf) properties are probed by a direct measurement of the temperature changes due to the absorption of radiation. The second system employs a point contact configuration, essentially to establish a very small two-dimensional constriction between two bulk metallic conductors. For this configuration, hf properties can simply be probed by measuring the radiation induced dc-voltages across the contact.

As a source of radiation in these experiments, an optically pumped far-infrared laser is used. Although a large number of discrete frequencies can be generated with it at relatively large power, it is not continuously tunable. Measurements therefore are preferably done at fixed frequency, but varying a second external parameter. For the present experiments, this external parameter was either a magnetic field or an electric field (voltage).

In both systems investigated, rather well known and conventional experiments are brought into a region where the prevalent conditions become qualitatively different. The Landau level separation becomes comparable to the Fermi-energy; the photon energy becomes comparable to the Fermi-energy, the superconducting energy-gap, the phonon energy or the plasma frequency. It is the purpose of the present work, to investigate how the phenomena, well known and understood under "classical" conditions, evolve under the above mentioned conditions. Although the sample materials used are very common, as the work is not intended to study particular materials, they are not completely arbitrary. They have been selected to emphasize, or make possible at all, the effects aimed at.

1. FIR magnetospectroscopy of semimetals.

A semimetal is an electrical conductor intermediate between a metal and a semiconductor. It has essentially all properties of a real metal, but the electron density is much lower and rather in the range typically found in degenerate (impure) semiconductors. The most familiar semimetal, also used in the present work is Bi. Because of its rather unique properties, very convenient for experiments it has been widely studied and has often served as a prototype metallic system for which new physical phenomena in metals have been discovered[1].

The metallic properties arise from a small overlap of a valence band at the T-point in the Brillouin zone with a conduction band at the L-point. The Fermi-energy is close to the bottom of the conduction band (30 meV) and to the top of the valence band. The number of electrons equals the number of holes ("compensated"). The exceptionally low band mass ($m^* \sim 0.01\ m_o$), in addition to the low Fermi-energy, provides the possibility to extremely quantize the electron system under application of modest magnetic fields. The extreme quantum limit, where one Landau level is left occupied below the Fermi-level is easily attainable.

The magnetoplasma effects in Bi have been investigated under conditions of extreme quantization at the corresponding FIR-frequencies. At lower frequencies, the magneto plasma effects are well known[2] and have been extensively studied.The electrodynamical properties are characterized by the occurrence of cyclotron resonances, hybrid resonances and dielectric anomalies. They can be analyzed and explained in terms of purely classical magnetoplasma theory, essentially based upon the Drude theory and the Lorentz force equation. At these frequencies (and magnetic fields) the behaviour of the solid state plasma in a semimetal is comparable to that for a gaseous plasma; the mass and velocity parameters are determined by their value at the Fermi level. This typical classical collective plasma behaviour has been investigated under conditions where the carrier system is strongly quantized. This is moreover of interest as the Bi bandstructure is known to be strongly nonparabolic, enhancing the influence of quantum effects.

The frequency range at which the experiments have been carried out links up with the microwave range at the low end and exceeds the plasma frequency at the high end. Over this range, all experimental results have been compared with calculations. Calculations have been carried out with the classical theory as well as with the quantum theory which uses a standard expression from time dependent perturbation theory, together with theoretical expressions for the velocity matrix elements applying to the relevant band-structure. The calculated results in the quantummechanical model do describe the data, which have been used as experimental tests of the expression for the matrix element. Many fine structure in the data, which is only present under quantum conditions, is reproduced by the calculations and forms a particular and previously unobserved (or unidentified) result of the non-parabolic bandstructure of Bi.

2. Point contacts and radiation.

The point contact configuration has a long history as a device used in high-frequency electronics. The so called cat-whisker detector diode was already employed in early crystal radios and is still familiar in microwave receivers. It consists of a sharpened, thin wire which is gently contacted with some base material. When the resulting current-voltage characteristic (I(V)) is nonlinear, the diode can be used for detection of radiation (by rectification) or more generally as a nonlinear element (e.g. mixer). The wire, which is part of the contact, acts hereby as a (free space) antenna to couple high-frequency currents across the contact when irradiated by an external hf field.

The metal-semiconductor or Schottky diode is the most frequently used combination in practice. A metal-metal contact, with a thin insulating oxide layer between the electrodes ("MIM") is used for very high frequencies (infrared and towards the visible light region). Point contacts between superconductors, which display the Josephson effect have been intensively investigated also, both for fundamental reasons and for their use in applications. Schottky diodes and MIM diodes can be used at room temperature, but Josephson junctions have to be operated at low temperatures. Josephson junctions and Schottky diodes can be used up to far-infrared frequencies (and even somewhat higher), but the operating range of MIM diodes is much larger[3].

Despite the poor mechanical stability and consequent limited reliability, the high frequency and broad band characteristics of point contact devices (notably MIM) are unequalled in any other device. For this reason, they are still a crucial element in accurate and absolute laserfrequency measurements, which are of great importance for laserspectroscopy[4a] and in establishing fundamental physical standards[4b].

The present work is concerned with the high-frequency limitation of the operation of classical, resistive (i.e. excluding the Josephson effect) nonlinear elements. The rectification behaviour of two particular point contact diodes is investigated under conditions where classical sensitivity approaches a quantum efficiency of unity. For this purpose, diodes have to be selected with very sharp nonlinearities and constriction type contacts (i.e. without an insulating layer) between either normal metals (NcN) or between a normal metal and a superconductor (NcS) have been chosen. In these contacts, which have a constriction diameter a small compared to the electronic mean free path l, current transfer is controlled by field emission processes. For useful results, operation at low temperature is required. The current-voltage characteristic for both types of diodes is by now very well understood theoretically on a microscopic level. For NcS contacts, the nonlinearity is governed by the process of conversion of supercurrent into normal current (Andreev reflection)[5] at the interface and occurs on an energy scale $eV \sim \Delta$ where e electron charge, V applied voltage and Δ half the superconducting energy gap. The nonlinearity in the NcN contact is due to an increased scattering rate when electrons are injected at energies near the zone boundary accoustic phonons[6] and occurs on an energy scale $eV \sim k\Theta_D$ (kO_D: Debye energy). The relevant energy scale in NcN contacts is therefore at least an order of magnitude larger than for NcS contacts and results in much weaker effects. Quantum effects are strongly present in the data for the NcS contacts, but also have been assigned to some anomalous features observed on NcN contacts at the highest frequencies used.

The experimental results have been analyzed with the so called Photon Assisted Tunneling model, which was originally developed for superconducting tunnel junctions[7] and is theoretically justified previously also for the current transport mechanism in NcS contacts[5]. The quantum effects described by this model have recently attracted a great deal of interest because of its implications for the use of superconducting tunnel junctions in practical microwave receivers[8]. In effect, this model states that the time varying voltage oscillations result in a steady state with quantized voltage amplitudes of separation $\hbar\omega/e$. It can be seen only when the dc I(V) characteristic is sharp on a voltage scale of $\hbar\omega/e$. The breakdown of the classical rectification model will also be discussed in terms of the electronic relaxation time concept ($\omega\tau > 1$).

The thesis is organized as follows. Chapter II provides some background concerning the experimental methods used and gives a description of the main parts of the experimental set-up. Data and analysis of the magnetospectroscopy of semimetals in high fields is given in Chapter III. Chapter IV contains a detailed analysis of rectification by point contacts at low temperatures and establishes a reference framework for Chapter V, which deals with the quantum effects in point contact rectification.

REFERENCES

1 . For a review, see: V.S. Edelman, Adv. Phys. 25, 555 (1976)

2 . G.E. Smith, L.C. Hebel and S.J. Buchsbaum, Phys. Rev. 129, 154 (1963)

3 . For a comparative review of different nonlinear elements, see: D.J.E. Knight and P.T. Woods, J. Phys. E: Sci. Instr. 9, 898 (1976)

4a. K.M. Evenson and F.R. Petersen in: *Laserspectroscopy of Atoms and Molecules* ed. by H. Walther (Springer, Berlin, 1976)

4b. H. Hellwig, K.M. Evenson, and D.J. Wineland, Physics Today 31, 23, dec. 1978

5 . S.N. Artemenko, A.F. Volkov and A.V. Zaitsev, Solid State Commun. 30, 771 (1979); A.V. Zaitsev, Zh. Eksp. Teor. Fiz. 78, 221 (1980) [Sov. Phys. JETP 51, 111 (1980)]; G.E. Blonder, M. Tinkham and T.M. Klapwijk, Phys. Rev. B 25, 4515 (1982)

6 . For a review see: A.G.M. Jansen, P. Wyder and F.M. Mueller, Science 199, 1037 (1978); A.G.M. Jansen, thesis, University of Nijmegen (1980)

7 . P.K. Tien and J.P. Gordon, Phys. Rev. 129, 647 (1963); J.R. Tucker and M.F. Millea, Appl. Phys. Lett. 33, 611 (1978)

8 . For a review, see: P.L. Richards and T.-M. Shen, IEEE Trans. Electron Devices ED-27, 1909 (1980)

CHAPTER II. EXPERIMENTAL METHODS.

1. Introduction.

In this chapter, the experimental'set up will be discussed. In addition, parameters governing the over-all performance of the system will be discussed in some detail, giving insight into design criteria and ultimate capabilities. For both types of experiments discussed in the next chapters, the heart of the set up was formed by the optically pumped far-infrared (OPFIR) laser. For doing spectroscopy on highly reflecting materials (i.e. metals), the technique of making the sample part of a cavity to increase the absorption is well established. Therefore, some effort was given to investigate Fabry-Perot types of cavities, suitable for the far infrared. Some results of this will be given and their potential use will be discussed. Finally, the absorption detection technique used for the experiments in chapter III is discussed. Detection occurred by a calorimetric method. Although perhaps not completely logical, it turns out to be convenient to arrange a discussion of the subparts in the following order. First, a rather general discussion of cavities and resonators is given, without too much reference to their use in the present experiment. Next, the laser performance and associated parameters are summarized. Finally, the calorimetric detection technique is outlined and the combined use of cavities is discussed.

2. FIR Fabry-Perot cavities.

The absorption of highly conducting opaque materials ((semi-)metals, perhaps also degenerate semiconductors) is very low. Usually, the reflectivity is the only measurable parameter. Because changes in absorption must be detected against the high background reflectivity, sensitivity is limited. This problem can be overcome by making the sample part of a cavity to increase absorption and decrease reflection. The type of cavities used successfully in the microwave region, cannot be used for the far-infrared.

For these cavities, dimensions are in the order of a wavelength so that fabrication becomes a difficult problem in the FIR ($\lambda \sim 0.1 - 1$ mm). A more fundamental limitation is however that reflection losses intrinsically are inhibitively high for obtaining useful high quality cavities of this type. Cavities (or resonators) used at optical frequencies of the Fabry-Perot (FP) type, are however very useful in the FIR. To fully exploit the potential merits of FIR FP cavities, the mirrors should be spherical. The spherical (or hemispherical) FP cavity has practical dimensions ($\sim$cm) with high quality factors. Dimensions are much smaller than for plane-parallel FP's, but much larger than single mode microwave types of cavities. Theoretical analysis (and so design considerations) of spherical resonant cavities is however of great mathematical complexity, but their basic performance is similar to a plane parallel FP, so that design and performance evaluation can be based on the plane parallel FP. To illustrate the definitions of the several practical parameters, some standard textbook formulas are summarized in the following.

2 i. Plane parallel Fabry-Perot.

Fig. 1 shows the plane parallel FP geometry. The upper mirror M_1 is semi-transparent with amplitude reflectivity r_{21}, power reflectivity R_1 ($=|r_{21}|^2$), amplitude transmissivity t, power transmissivity T ($=|t|^2$), power absorptivity A_1. Although not relevant for the analysis, we have assumed the second mirror to be opaque, which would correspond to the experimental configuration where the sample acts as mirror M_2. Mirror M_2 has amplitude and power reflectivity r_{23} and R_2 ($R_2=|r_{23}|^2$) respectively and absorptivity A. $1=T+R_1+A_1$ and $1=R_2+A$. Mirror separation is d and incident radiation has amplitude E_0 and intensity I_0. The refractive index of the medium between the mirrors is n (practically n=1). With the usual addition of amplitudes, $E_1+E_2+\ldots$, one obtains for the circulating intensity I between the mirrors the appropriate

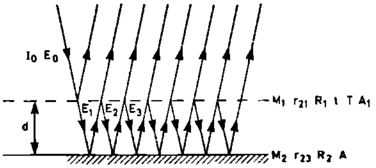


Fig. 1. Schematic illustration of a plane-parallel Fabry Perot configuration.

version of the Airy formula:

$$\frac{I}{I_0} = \frac{T}{(1-\sqrt{R_1R_2})^2} \{1 + \frac{4\sqrt{R_1R_2}\sin^2(\alpha/2)}{(1-\sqrt{R_1R_2})^2}\}^{-1} \tag{1a}$$

with the phase parameter α for the relevant case of normal incidence given by:

$$\alpha = 4\pi nd/\lambda + \phi \tag{1b}$$

where ϕ is a constant phase factor, containing the phase of $r_{23}r_{21}$ and λ the vacuüm wavelength. (For incidence at an angle i with respect to the normal to the mirrors, the first term of (1b) is multiplied by cos(i)).

The conventional form of the Airy formula, giving the Fabry Perot transmission is obtained from (1a) by multiplying with T_2 if T_2 were the transmissivity of M_2. We are interested in the absorptivity A_{FP} with respect to the incoming beam I_0 when mirror M_2 (the "sample") is part of the FP cavity. A_{FP} is found from (1a) by multiplying with A.

(1a) describes the familiar sharply peaked FP interference pattern as a function of α. The cavity resonates for $\alpha=q.2\pi$ with q integer. From (1) it follows that the pattern is periodic, with periodicity depending on the ratio nd/λ and given by $\Delta(nd/\lambda)$:

$$\Delta(nd/\lambda\ ) = 1/2 \tag{2}$$

The half-widths of the peaks (full width at half maximum, FWHM) as a function of nd/λ, $\delta(nd/\lambda)$ follows from (1) as:

$$\delta(nd/\lambda) = (1/\pi)\arcsin((1-R)/2\sqrt{R}) \tag{3}$$

with R defined as $R = \sqrt{R_1R_2}$. (For definitions, see Fig. 2).

A useful parameter for a FP is the reflective Finesse F_R defined by $F_R = \Delta/\delta$. Usually, $1-R \ll 1$, so that from (2) and (3)

$$F_R \simeq \pi\sqrt{R}/(1-R) \simeq \pi/(1-R) \tag{4}$$

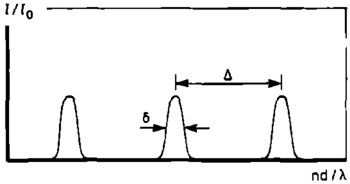


Fig. 2. Schematic illustration of a Fabry Perot performance pattern, showing the definitions of δ and Δ.

Experimentally, F_R is directly found from mirror translation scans at fixed frequency. The frequency periodicity $\Delta\nu$ of a resonator of fixed length is called the free spectral range (FSR). From (2):

$$FSR = c/2nd \tag{5}$$

with c speed of light. If a (transmitting) cavity is illuminated by broad band radiation, only discrete frequencies, separated by FSR, are excited and transmitted. It is simple to show that in the time domain a single pulse is then bouncing back and forth between the mirrors, so that a pulse train leaves the cavity, the pulses separated in time by the cavity round trip time $\tau_{rtt} = 2nd/c = 1/FSR$. Single pulse width is given by $1/(FSR.m)$ with m the number of discrete frequencies. Such behaviour requires phase cohaerence of the discrete frequencies and can occur in lasers ("mode locking").

The frequency resolution $\delta\nu$ of a cavity of fixed distance is given by the frequency width of the peaks for fixed d (FWHM). From (3), with the definition of F_R and $Q \equiv \nu/\delta\nu$:

$$Q = \nu/\delta\nu = \nu/FSR\ .\ F_R = (2nd/\lambda).F_R \simeq q.F_R \tag{6}$$

Q is the quality factor and q the number of half wavelengths in the resonator. With these definitions, the peak shape of a resonance either as a function of cavity distance at fixed frequency, or as a function of frequency at fixed cavity distance is usually written for the two cases as:

$$I/I_0 = (I/I_0)_{max} \cdot \{1+4F^2((d-d_0)/(\lambda_0/2))^2\}^{-1} \tag{1a'}$$

with d_0 the distance at resonance and λ_0 the fixed wavelength, or as:

$$I/I_0 = (I/I_0)_{max}.\{1+4Q^2((\omega-\omega_0)/\omega_0)^2\}^{-1} \qquad (1a'')$$

with ω_0 the resonant frequency. From the equivalent definition of Q, $Q = \omega I/(dI/dt)$ (I intensity in the cavity), the 1/e decay time of the radiation in the cavity (or lifetime of a photon in the cavity), τ_c is:

$$\tau_c = Q/\omega = 1/2\pi\delta\nu \qquad (7)$$

The total number of roundtrips before decaying is thus given by τ_c/round trip time:

$$\text{number of roundtrips} = F_R/2\pi \qquad (8)$$

2 ii. Hemispherical Fabry-Perot.

So far, diffraction effects were neglected. The natural angular beam spread, due to diffraction is $\theta\sim\lambda/an$, where a is the radius of the beam. For a standing wave in a plane parallel Fabry Perot, the maximum beam radius is of the order of the mirror radius (or linear dimension of the mirror). The number of roundtrips the radiation makes before "walking out" between the mirrors is thus maximally given by $a/\theta d = a^2/\lambda d$. The quantity $a^2/\lambda dn$ is known as the Fresnel number and is a measure of the diffraction losses ($1/N$ = fractional energy loss per transit, roughly). In order to have negligible diffraction losses compared to the mirror reflection losses we have (see (8)):

$$N >> F_R \qquad (9)$$

It is precisely relation (9) which makes a plane parallel FP unsuited for the FIR, unless mirror dimensions are large. The use of spherical mirrors with a periodic focussing action greatly reduces the diffraction losses. Moreover, and especially important for the present purposes, a (hemi-) spherical Fabry Perot does not require a severe alignment. The hemispherical cavity is the common type of (gas)laser cavity and is thoroughly analysed[1,2]. One way to derive the resonant conditions and modes of the (hemi)spherical resonator is analogous to the method for the plane parallel, that is starting with the free space solutions of the Maxwell equation $\nabla^2E+k^2E = 0$. The set of plane wave solutions however is now not very well suited because the finite beam width and associated diffraction is essential. Instead, the Gaussian beam solutions are used[2]. They are found by inserting solutions of the type $E=\psi(x,y,z)e^{-ikz}$ in the wave equation[2]. They form a complete and orthogonal set of functions and the eigenfunctions are designated as modes. All modes are characterized by the complex beam parameter q given by:

$$1/q = 1/R - i\lambda/\pi w^2 \tag{10}$$

The amplitude of each mode is modulated transverse to the propagation direction by a Gaussian $\exp-((x^2+y^2)/w^2)$ and the beams have spherical constant phase surfaces of radius of curvature R. R and w are given by[2]:

$$R(z) = z\{1+(\pi w_0{}^2/\lambda z)^2\} \tag{11a}$$

$$w^2(z) = w_0{}^2\{1+(\lambda z/\pi w_0{}^2)^2\} \tag{11b}$$

w_0 is a constant, giving the minimum beam radius as a function of z (beam waist). The amplitude distribution below the Gaussian envelope is characterized by 2 mode numbers (r,s) and explicit expressions can be found in ref. 2 or standard textbooks.

Resonant modes of a hemispherical resonator are found as the free space Gaussian beam solutions which repeat themselves after each round trip, which can be shown to imply that the beam curvature (11a) matches the mirror curvature. Condition for resonance is now:

$$d = \frac{\lambda}{2}\,\{q+(r+s+1).(1/\pi).\arccos\sqrt{(1-d/R_1)(1-d/R_2)}\} \tag{12}$$

with q=1,2,..., r,s = 0,1,2,... ,R_1 and R_2 the curvature radii of the two mirrors. The beam width (radius) at the position of the two mirrors for the hemispherical case ($R_2=\infty$) is given by:

$$w_1{}^4 = (R_1{}^2\lambda^2/\pi^2).d/(R_1-d) \tag{13a}$$

$$w_2{}^4 = (\lambda/\pi)^2\ .\ d(R_1-d) \tag{13b}$$

When $d\to R_1$, the spot size at the curved mirror becomes arbitrarily large and that at the flat mirror small. The modes are designated as TEM_{rsq} or TEM_{rs}.

The first term of (12) can be seen to be the equivalent of the plane parallel FP resonance condition and q is the longitudinal mode number. The "fundamental" longitudinal modes TEM_{00q} have a purely Gaussian intensity distribution transverse to the beam axis. The "transverse" or "off-axial" modes TEM_{rsq}, $r,s \neq 0$ have a modulated intensity profile transverse to the axis. They are the equivalent of the familiar ring-shaped interference pattern of a plane parallel FP when illuminated by a beam of finite angular spread. The "rings" however are not pure modes (12), but combinations of degenerate modes or syntheses from different polarization configurations. In fact, a (single) ring shaped intensity pattern emanating from a (gas)laser cavity frequently occurs and forms one of the most notorious laser modes ("doughnut"). From (12), it is seen that the off-axis modes increase the number of possible resonances. Formally, taking the off-axis resonances into account, the definitions of Finesse and free spectral range should be modified accordingly (and is sometimes done). It is however useful to explicitly distinguish between axial and off-axial modes.

A cavity is called stable when the radiation does not walk out between the mirrors, i.e. that periodic refocussing is effective. The stability criterion for a spherical cavity is:

$$0<(1-d/R_1)(1-d/R_2)<1 \tag{14}$$

The losses for the spherical FP due to the finite mirror dimensions have been calculated also[1,2]. The Fresnel number N ($=a_1a_2/\lambda d$ with a_1, a_2 the radii of the mirrors) is still a measure for the diffraction loss. The losses depend further on the product $(1-d/R_1)(1-d/R_2)$ and are mode dependent. For $N<<1$, the losses are as large as for a plane parallel FP, but for $N\underset{\sim}{>}>1$ are orders of magnitude smaller. Lowest losses exist for the confocal resonator ($d=R_1=R_2$ or $R_2=\infty$, $d=\frac{1}{2}R$). For the hemispherical FP ($R_2=\infty$) or the double spherical FP (with $R_1=R_2$), the diffraction losses approach those of a plane parallel FP for $\frac{1}{2}R << d < R$ and $R<< d < 2R$ respectively. Diffraction losses rapidly increase with higher order modes, but this is the case for the plane parallel FP as well. For $d << R$, the losses are comparable to those for the plane parallel FP with the same Fresnel number. Quantitative calculations can be found in refs. 1 and 2. Table I gives some typical values for the fractional diffraction loss per pass α as obtained from the plot in ref. 2 for typical cavity parameters which might be used in an experiment (see section 5 of this chapter). The losses for plane

λ	α plane parallel	α spherical
0.10 mm	0.01	$<<$ 0.002
0.75 mm	0.12	0.01

Table I. Comparison of losses for different FP cavities. $a_1a_2/\lambda d \sim 12$ *and 1.6 for* $\lambda = 0.1$ *and0.75 mm respectively;* $d \sim 17$ *mm,* $R \sim 19$ *mm.*

parallel and spherical cavities are compared, assuming both have the same Fresnel number For the hemispherical cavity the losses are strongly determined by $|1-2d/R|$. The diffraction losses should be compared to the reflection (and absorption) losses, which are a few percent. The improvement of the spherical cavity over the plane parallel is clear from Table I. It may be seen that α spherical is even somewhat large at 0.75 mm and would be unacceptibly high at longer wavelengths. However, α spherical is very strongly dependent on N as well as on $|1-2d/R|$ in this region so that the behaviour can be best tested in practice. Lower values are obtained by decreasing d (but this is to some extent limited in practice) or increasing R.

2 iii. Experimental tests of cavity configurations.

To test the performance of a hemispherical FP in practice, some simple, room temperature test experiments have been carried out. For this purpose, FP configurations have been realized which are compatible with a possible use in the low temperature insert. Because absorption is not easily measurable at room temperature, only transmission type FP 's have been used for the tests. This will be a valid check of their performance as far as the geometry is concerned. One of the mirrors was mounted on a precision translation stage which could be moved by a motor-driven micrometer. The flat mirror was always a free standing rectangular metal mesh (inductive)[3]. The mesh constant was chosen to give nominally high reflection (>95%) at the wavelength of interest. Typical absorptivities of such meshes are in the order of 0.5%[3]. Because of the small mirror diameter, the mesh surface was plane when the mesh was glued on a ring, without additional stretching. Two different types of spherical mirrors have been used, following earlier work of Allen et al.[4] and Goy et al.[5] respectively. In the first[4], a metal mesh was glued against the concave surface of a single crystal quartz lens, using diluted General Electric 7031 (GE) varnish. (see Fig. 22 in section 5). Although the mesh is necessarily deformed in this way, a reasonable smooth surface could still be obtained. The commercially obtained quartz lens had an inner radius of curvature R = 19 mm. The lens was mounted at the end of a lightpipe, which was illuminated with the FIR radiation.

In the second type, the spherical mirror is machined from copper and hole input coupling is employed. This kind of approach is typical for the microwave region, either for rectangular cavities,or also hemispherical microwave FP 's^{6}. It follows the design of Goy et al.^{5}. A cone of low angle ends up in a short (few mm) piece of cylindrical waveguide, with dimensions of the order of a wavelength, so that a single low order (fundamental) mode may be excited in the waveguide, see Fig. 3. The coupling diaphragm should be as thin as possible (<<λ). The optimum radius of the coupling hole depends on the losses in the cavity (see section 5). If these are small, the radius should be smaller than λ. For radii >>λ, reflection losses through the hole become very important. The actual dimensions are somewhat determined by practical possibilities. The cone had a top angle of ∿6° and waveguide diameter ∿0.6 mm, and both were made by copper plating machined and polished aluminum. The spherical mirror had a hole of ∿0.3 mm diameter, first drilled into the copper block, after which the curved mirror was machined leaving the final diaphragm. With skillful machining on a high accuracy apparatus, thicknesses in the order of 0.1-0.3 mm could be obtained with a negligibly small gap between the waveguide end and the diaphragm. The final mirror surface, which had a radius of curvature of 4 cm, was hand-polished. The (in plane) radius of the spherical mirror was 6 mm, that of the flat mirror ∿5 mm. The tests were done using the FIR-laser, to be described in the next section, as a monochromatic source.

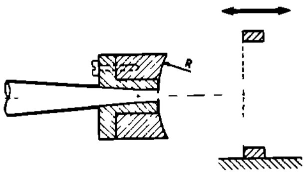


Fig. 3. Hemispherical Fabry Perot cavity with hole coupling. Spherical mirror, cone and lightpipe are from copper. For dimensions and realization: see text. The flat mirror is a free standing copper mesh.

Fig. 4 shows a typical behaviour obtained with the curved mesh mirror. The mesh constant g is 100 μm and λ=1.22 mm so that λ>>g. The three spectra correspond to three different mirror spacings. The periodicity of the main peaks corresponds to the excitation of fundamental TEM$_{00}$ modes. The additio-

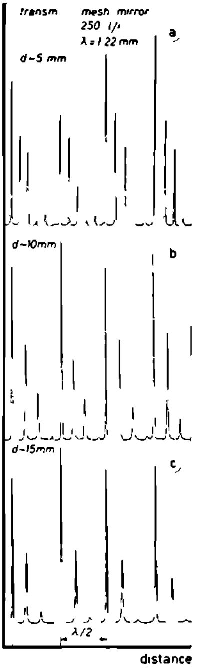


Fig. 4. Hemispherical Fabry Perot transmission scans at three different initial mirror separations.

nal peaks between the fundamental modes are higher order transverse modes TEM_{rs}. It can be seen that the distance between transverse modes depends on the distance between the mirrors, especially at small distances. The separation between secondary modes increases with distance in run (a), made with very short mirror distance and increases upon going from (a) to (c). From (12), taking $R_2=\infty$, one finds for the separation Δ_{tm} (in length units) between adjacent transverse modes:

$$\Delta_{tm} = \frac{\lambda}{2}\cdot\frac{1}{\pi}\cdot\arccos\sqrt{1-d/R_1} \qquad (15)$$

which follows from (12) when changing r+s by unity, leaving q unchanged. Measuring Δ_{tm} from the data gives an independent way to calculate d. In this way, one finds that in (a) d changes from 5 mm to 7 mm, corresponding to three $\lambda/2$ periods. The distances for curves (b) and (c) are found as 11 and 13 mm respectively. These values for d are in very reasonable agreement with the directly (but not very accurately) measured values of d. It shows that the assignment of the secondary peaks as adjacent TEM_{rs} modes belonging to the same longitudinal mode number q is correct.

In principle, all possible modes are excited, apart from degeneracy; that is, a peak is present at each position Δ_{tm}. However, the amplitude is strongly mode dependent, decreasing rapidly upon further distance from the fundamental. In (a), modes with $r+s=1,2$ are strong, but modes with $r+s>2$ are barely visible. This indicates again that the higher order modes belong to the same q number as the immediately preceeding fundamental mode. Although the amplitude of an excited mode does depend upon the (unknown) field distribution outside the cavity, the rapid decay with increasing transverse mode numbers is caused predominantly by the increased diffraction losses[1,2]. The effective Finesse from Fig. 4 is found as $F_R \simeq 60$, corresponding to a mirror reflectivity $R = 0.95$ (eq. 4) if Finesse was determined by reflection losses only. The mirror reflectivity cannot be accurately determined, especially also because one of the meshes is glued on a dielectric so that its reflectivity depends also on the substrate parameters[3] (quartz, GE-varnish). A value of $R=0.95$ however seems somewhat low compared to the high wavelength to mesh-parameter ratio λ/g. A value of $F_R \sim 60$ however is still very satisfactory. The Fresnel number for this cavity at the wavelength used, varies from $N=5$ at $d=5$ mm to $N=1.6$ at $d=15$ mm. The maximum Gaussian beam width at the curved mirror (at $d=15$ mm) is calculated from (13a) as $2w_1=7.6$ mm, much less than the mirror diameter. Consistent with calculations of diffraction losses[1,2], it can be concluded that very good cavity performance is obtained, even at such very low Fresnel numbers where a plane parallel Fabry Perot would not work at all. It is moreover interesting to note and useful for practice that the behaviour shown in Fig. 4 was obtained without any alignment of the mirrors, apart from eye-setting them in the optimal position.

In Fig. 5 two spectra are compared, obtained with the copper mirror and the mesh mirror at $\lambda=571$ µm. With the mesh mirror a Finesse $F_R=36$ is obtained about half the value obtained at $\lambda=1.22$ mm with the same mesh. With the known wavelength dependence of the mesh reflectivity, one finds theoretically[3] $F_R \sim \lambda^2$, whereas the data are closer to linear. This might indicate that the Finesse at 1.22 mm was limited by diffraction losses (possibly enhanced due to mirror misalignment). With the copper mirror and wavelength close to the nominal optimal design wavelength $F_R \simeq 85$. This high value for F_R indicates a very high effective reflectivity for the copper mirror (very near 1). The reflectivity of the mesh can be estimated from Fig. 5b with $F_R=36$: $R_{mesh} \sim 0.92$. Because F_R in Fig. 5a is determined by $\sqrt{R_1 R_2}$ (eq. 1a) one finds for the reflectivity of the copper mirror $R_{copper}=1.00 \pm .01$, that is indistinguishable

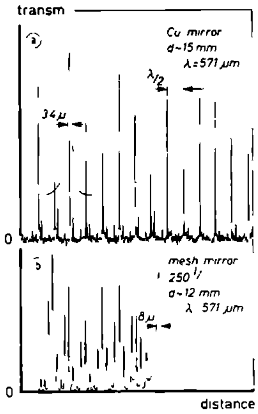


Fig. 5. Comparison of two FP transmission scans obtained with the spherical copper and mesh mirrors. The dimensions for the copper mirror configuration are near optimum for the wavelength used and the flat mirror employs 250 lines/inch mesh.

from unity in combination with the mesh reflectivity. The losses through the coupling hole are thus comparable to, or perhaps much smaller than the reflection losses from the copper surface and so satisfying the design criteria.

Clearly, with the second mirror also of high reflectivity (e.g. also a solid polished copper mirror), cavities of the type of Fig. 3 can have very high F_R values and extremely high Q 's. This property also is well known from their upscaled versions at microwave frequencies[6].

It is not evident that there will be more modes with the mesh mirror than with the copper mirror. The parameter $|1-2d/R|$, determining the diffraction losses is nearly the same in both cases ($|0.25|$, although of different sign). The Fresnel numbers differ somewhat (3.5 for the copper mirror and 4.4 for the mesh mirror) and although the diffraction losses depend very strongly on Fresnel number in this range[1,2], the losses should still be negligibly low for the first several higher order modes in both cases. Presumably, alignment effects are of importance here, which obviously are more important for higher order than for fundamental modes. A discussion of the peak amplitudes will be left for section 5 of this chapter, but it may already be noted that the cavity of Fig. 5a is heavily undercoupled (coupling losses << absorption and other transmission losses).

Fig. 6 shows a scan with the copper mirror at a wavelength smaller than the coupling hole diameter. The flat mirror for this wavelength was made from finer mesh (mesh constant 25 μm). From FP scans with the spherical mesh mirror (not shown), values $F_R \sim 50$ were obtained for this wavelength and mesh combination. From Fig. 6 a Finesse $F_R \sim 42$ is obtained. Therefore, in contrast to

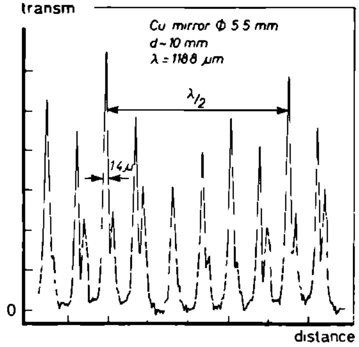


Fig. 6. Hemispherical FP transmission scan with the copper mirror FP at a wavelength much smaller than optimum. Flat mirror: 1000 lines/inch mesh.

the situation discussed for Fig. 5, it now seems that the replacement of the (even not too high reflecting) mesh by the copper mirror does not improve the effective Finesse. This means that coupling losses through the hole are now considerable, as expected for a hole larger than the wavelength. Moreover, the coupling losses are presumably now better matched to the other losses in the cavity (as compared to the case of Fig. 5, see also section 5).

An extra absorbing diaphragm (diameter 5.5 mm) was introduced in the cavity, intended to reduce the higher order modes. It has been verified that this works very well in practice but in case of Fig. 6 the higher order modes are still of considerable strength. This is consistent with the high Fresnel number $N \simeq 11$ with the area of one mirror determined by the diaphragm. The high N-value suggests also that the peak widths in Fig. 6 are not limited by diffraction losses. All modes are present, whereever expected and from their distance one finds, using (15) and the nominal value $R_1=40$ mm, $d \simeq 10$ mm, in good agreement with the nominal setting of the mirror distance. It is not clear why some of the modes in Fig. 6 seem splitted. Presumably this splitting is related to alignment effects.

A long FP scan at large mirror distance is shown in Fig. 7. With an initial nominal mirror separation of 17 mm and the radius of curvature of the mirror R=19 mm, the cavity is expected to become unstable after 7 half wave lengths ($d \sim R_1$). It can be seen from the figure that the initial sharp FP peaks gradually become broader for d approaching R and the typical FP spectrum gradually decays for d>>R, in the unstable region. For $d \gtrsim R$, the peaks seem assymmetric, consisting of a rather sharp rise and a slow decay; peak width is in the order of the peak distance. A similar picture could not be obtained with the R=40 mm radius of curvature mirror. Presumably diffraction

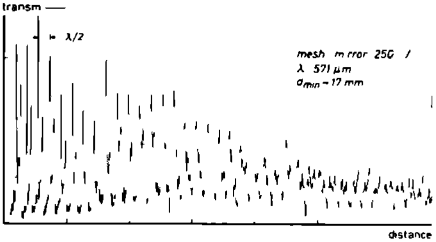


Fig. 7. Extended FP transmission scan at large initial mirror separation. The FP pattern disappears as the cavity becomes unstable at large mirror distances (d>R).

losses became too high before the unstable region was reached due to the low Fresnel number. The FP spectra decayed at distance d<R. The unstable region is not of interest for the present applications, but unstable resonators have some useful (laser)applications (higher mode volume).

Concluding this section, it can be stated that the small hemispherical Fabry Perot is extremely simple and easy to use in the submillimetre region. The Fabry Perot performance itself is well described by the theory. Efficient excitation (incoupling) of power is however a very difficult problem in this frequency range and will be discussed further in section 5.

3. The optically pumped far-infrared (OPFIR) laser.

The considerable development of the optically pumped far-infrared laser since its first demonstration by Chang and Bridges[7], has made it by now an increasingly more important source of radiation in the submillimetre region. Its basic operation principles and practical realization have been described in several reviews[8-11]. A schematic outline of the system used in the present work is given in Fig. 8, together with the arrangement for the experiments of chapter IV and V.

3 i. The pump (CO_2) laser.

Key part of the system is the CO_2 laser[12], to pump the FIR laser. Laser action occurs on the vibration-rotation levels of the CO_2 molecule (Fig. 9). The upper laser level is the lowest excited assymmetric stretch mode (001) of the linear O-C-O molecule. The lower laser level can either

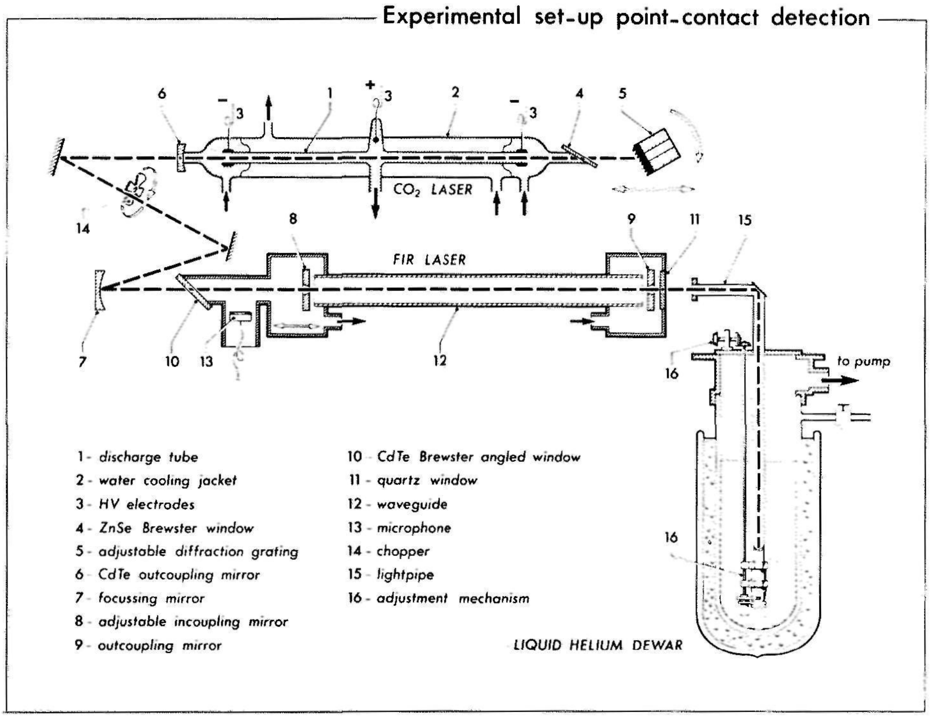


Fig. 8. Schematic outline of the optically pumped far infrared laser system, together with a typical experimental arrangement used for the work described in chapters IV and V

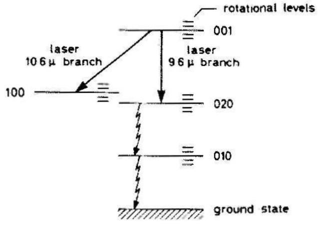


Fig. 9. Relevant part of the energy level diagram of the CO_2 molecule.

be the second excited bending mode (020) (with emission wavelength in the 9.6 µm region) or the lowest excited symmetric stretch mode (100) (with emission wavelength in the 10.6 µm region). Each vibrational level is splitted in unequally spaced rotational levels, so that each vibrational transition is splitted in a large number of frequencies. The rotational states are described by one quantum number J only and selection rules are $\Delta J=\pm 1$. Transitions with $\Delta J=+1$ establish a long wavelength branche (P branche), those with $\Delta J=-1$ a short wavelength branche (R branche). The transitions are conventionally designated as 9RJ, 9PJ, 10RJ,10PJ where the 9 and 10 denotes the 9.6 or 10.6 vibrational transition bands respectively.

Excitation of the upper level occurs in a gas discharge. The useful properties of the CO_2 laser depend however essentially on the additional gases N_2 and He. In the discharge a long lived vibrational N_2 state is effectively excited, which by collision processes selectively and resonantly excites the upper CO_2 laser level. Selective excitation is crucial for efficient laser operation. The He gas in the discharge is believed to increase the electron density in the discharge and thereby the excitation of the N_2 and CO_2 molecules. It seems to play an important role in the depletion of the lower level to the ground state also.

3 ii. The FIR laser.

The relevant molecular energy level scheme applying to the FIR-laser active molecules is shown in Fig. 10. The two lowest vibrational levels of the molecule are shown which are splitted in rotational levels. If one of the CO_2 laser lines accidentally coïncides with a particular set of levels of the

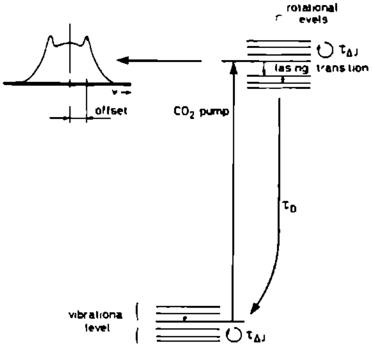


Fig. 10. Relevant molecular energy level scheme applying to FIR-laser active molecules. Important relaxation processes are schematically indicated. The inset shows the pumped population distribution of the upper FIR level.

FIR molecule, strong absorption occurs which may result in population inversion between a set of two rotational levels. The linewidth of the pumped transition for this purpose must overlap with the tuning range of the CO_2 laser. The laser properties are determined by the strength of the population inversion which can be maintained between the two laser levels. For the FIR laser, a crucial parameter is the collision rate which determines the thermalization of the rotational levels, which decreases the inversion, $\tau_{\Delta J}^{-1}$. The vibrational relaxation rate τ_D^{-1} counteracts the thermalization. The ratio $\tau_D/\tau_{\Delta J}$ should therefore be as small as possible. The threshold condition for obtaining positive gain has been stated by Tucker as[10]:

$$\frac{h\nu_{FIR}}{kT} f_J \frac{\tau_D}{\tau_{\Delta J}} < 1 \tag{16}$$

where $h\nu_{FIR}$ the FIR photon energy, kT Boltzman constant times temperature and f_J the Boltzman factor for the upper rotational state. $\tau_{\Delta J}^{-1}$ is proportional to the pressure. Vibrational relaxation is a very slow proces and is in practice dominated by diffusion out of the amplifying region or through collisions with the wall[10]: $\tau_D \propto pD^2$ where p the pressure and D the diameter of the active laser region (∿diameter of the waveguide for a waveguide laser). From (16) it then follows that the threshold is determined by the ratio $\tau_D/\tau_{\Delta J}$ or by the experimentally controllable parameter p^2D^2. From the threshold criterion (16), a critical pressure p_c can be defined above which no gain can be achieved, $p_c \propto D^{-1}$.

Although the relative inversion increases with decreasing pressure, FIR power output is evidently proportional to the absolute inversion, which increases with pressure. Moreover and most important, at low pressures the pump beam is not absorbed, leading to low efficiencies. As a result of these competing effects, there is a relatively sharply defined optimum pressure for FIR laser operation, in the order of the critical pressure and typically some few 100 mTorr (depending on the dimensions of the laser). The relation

$$p_{optimum} \propto D^{-1} \tag{17}$$

has been verified experimentally[13]. The low operating pressure of the FIR laser, with its awkward consequences is thus essentially due to the vibrational relaxation bottleneck. A mechanism which decreases τ_D, or makes τ_D^{-1} proportional to pressure, would greatly improve and facilitate FIR laser performance, but has yet to be found.

The width of the relevant levels also is of crucial importance for the laser operation. For gaslasers it is generally determined by collisional broadening and Doppler broadening. The dominant type of broadening is very important, especially in connection with the OPFIR laser, because collisional broadening leads to a homogeneous linewidth (each molecule contributes to the complete linewidth), while Doppler broadening results in an inhomogeneous linewidth (molecules with different velocities contribute to different parts of the lineshape). Collisional broadening is independent of frequency, but Doppler broadening is proportional to the frequency. At typical operating pressures, the CO_2 vibrational transitions are collisional broadened (in the order of $\gtrsim$ 100 MHz, the Doppler width being some 50 MHz for a cooled plasma). Because of the low pressure of the FIR gas, the CO_2 absorption transition is Doppler broadened ($\sim$50 MHz). At the low FIR frequencies however, the rotational transitions are predominantly collisional broadened, although mixed broadening may become important at the higher FIR frequencies. The FIR collisional broadened linewidth is given by $\Delta\nu \simeq 1/\pi\tau_{\Delta J}$ and is known to be in the order of 40 MHz/Torr.

3 iii. Operation principles OPFIR laser.

From the point of view of the laser theory, the physical processes are lumped into some convenient parameters to obtain a useful expression for the gain through stimulated emission:

$$g \equiv \frac{1}{I}\frac{dI}{dz} = \frac{g_0}{1+I/I_s} \qquad (18)$$

which can be obtained from rate equation models for a homogeneously broadened gaslaser. I is the intensity of the radiation field, z distance along the propagation and g_0 and I_s are constants : the small signal or unsaturated gain coefficient and the laser saturation intensity respectively. (The dependence of g on I for large intensities is obvious because the gain is proportional to the inversion, but the inversion depends on the rate of stimulated emission which is proportional to I). Gain saturation effects are important both for the CO_2- as well as for the FIR laser, as both are high gain systems.

With the amplifying medium contained in a suitable cavity to provide optical feedback, with partly transparent mirrors, (18) can be integrated to yield the familiar expression for laser output intensity of a homogeneously broadened laser:

$$I_{out} = \tfrac{1}{2}\, T\, I_s \left(\frac{2g_0 L_a}{\alpha} - 1\right) \tag{19}$$

T is the transmissivity of the outcoupling mirror, L_a the length of the active medium (usually $L_a \lesssim d$, d cavity length), α the total loss per pass, $\alpha \simeq T+A$ with A the absorptivity of the mirrors.

The saturation intensity is given by[14]:

$$I_s = h\nu_{FIR} c \cdot \frac{1}{1+g_2/g_1} \cdot \frac{1}{B\tau_{\Delta J}} \quad (\propto p^2) \tag{20}$$

which is a fairly general expression. c is the speed of light, $g_1 \sim g_2$ the degeneracy factors of the laser levels and B the Einstein coefficient for stimulated emission.

$$B = \frac{c^3\, f(\nu)}{4\pi^2 \nu^2 \tau_{sp} \Delta\nu_J} \tag{21}$$

with $f(\nu)$ the normalized Lorentzian of half width $\Delta\nu = (\pi\tau_{\Delta J})^{-1}$ and τ_{sp} the spontaneous lifetime.

For the optically pumped laser, the small signal gain ($\sim$population inversion) is, using a simple 3 level rate equation model, easily expressed in terms of the absorbed intensity of the CO_2 laser, I_{abs}[14]:

$$g_0 = B\, \tau_{\Delta J} \cdot \frac{I_{abs}}{h\nu_{IR} c L} \tag{22}$$

If γ is the absorption coefficient of IR radiation for the pumped transition and α_{IR} the cavity losses per pass at the IR frequency (mirror losses, cavity walk out, sidewall absorption etc.), then I_{abs} is found from the ratio of pump-transition absorption per pass/total loss per pass as:

$$I_{abs} = \frac{\gamma L}{\gamma L + \alpha_{IR}} \cdot \frac{I_{IR}}{h\nu_{IR} c L} \tag{23}$$

which holds for $\gamma L \ll 1$. L is the FIR cavity length and I_{IR} the pump intensity. Note that by definition of the several quantities, I_{IR} refers to the incoming pump beam, not to the circulating pump intensity. When IR cavity resonances are important, it is assumed that the cavity is tuned at resonance for the IR frequency also. Under conditions of high gain, (i.e. for optimum pressure), usually applying to a FIR laser, the FIR output intensity follows from (19), neglecting the factor 1 in the term between brackets[10]:

$$I_{out} = \left(\frac{1}{1+g_2/g_1} \cdot \frac{\nu_{FIR}}{\nu_{IR}}\right) \cdot \frac{T}{T+A} \cdot \frac{\gamma L}{\gamma L+\alpha_{IR}} \cdot I_{IR} \qquad (24)$$

The power output is found by multipying I_{out} by the beam diameter. The theoretical maximum power output is limited by the quantum conversion efficiency $(1+g_2/g_1)^{-1}(\nu_{FIR}/\nu_{IR})$ and would be obtained when the absorptive cavity losses at the FIR frequency, A and the total (passive) cavity losses at the IR frequency, α_{IR} were negligible.

The pump absorption coefficient γ depends strongly on pressure. For the typical low pressures employed, absorption saturation effects are important (maximally half the total number of molecules can be excited, so that the absorption coefficient strongly decreases with intensity at high intensities). The saturated absorption coefficient for an inhomogeneously broadened absorption transition is given by:

$$\gamma = \gamma_0/\sqrt{1+I_{IR}/I_{s,A}} \qquad (25)$$

where γ_0 is the small signal absorption coefficient and $I_{s,A}$ the absorption saturation intensity at the pump frequency. The expression for saturated absorption (25) is analogous to the expression for saturated (laser)gain (18), the square root arising from the fact that inhomogeneous broadening is assumed in (25). $I_{s,A}$ is given also by expression (20) however with the parameters applying to the pump transition of the FIR molecule. $\tau_{\Delta J}$ is to be replaced by the lifetime of the upper and lower pumplevels.

For most practical systems, the relation $\gamma L \ll \alpha_{IR}$ is difficult to avoid, so that I_{abs} is proportional to γ (see (22) and (23)). Because $I_{s,A} \propto p^2$ ((20)), the absorption is saturated at low pressures and $\gamma \propto p^2$ ((25), $\gamma_0 \propto p$). Although the FIR laser gain intrinsically varies as p^{-2} via $B\tau$ ((22)) the practical circumstance $\gamma L < \alpha_{IR}$ renders g_0 independent of pressure at low pressure by the pressure dependence of I_{abs}. In this situation, the power output increases with pressure via I_s. At higher pressures and/or relatively less cavity pump losses, I_{abs} increases less fast than p^2 (ideally, it is independent of p for low enough α_{IR}), can no longer cancel the intrinsic p^{-2} decrease of g_0 (via $B\tau$, (22)) and g_0 decreases to go below threshold ((19), (16)), the optimum pressure, for which (24) roughly holds, being determined by vibrational relaxation (16) and scaling with active medium diameter according to (17).

From (24), it is seen that the power output varies linearly with the pump power well above threshold (determined by $2g_0L_a=\alpha_{FIR}$, (19)),for not too strong absorption saturation. In the limit of strong absorption saturation (with $\gamma \propto I_{IR}^{-\frac{1}{2}}$, and $\gamma L << \alpha_{IR}$), $I_{out} \sim I_{IR}^{\frac{1}{2}}$. If the absorption transition was taken homogeneously broadened, i.e. omitting the square root from (25), which has apparently been done in ref.(14), the output power would have been expected to saturate at a constant level. An $I_{IR}^{\frac{1}{2}}$ dependence is predicted however in the more sophisticated model of ref. (15), where the pump transition was explicitly assumed to be inhomogeneously broadened. Sublinear dependence of I_{out} on I_{IR} has been observed well above threshold for pressures much below optimum[14], which was attributed to pump saturation effects, but at the optimum pressure a linear variation is invariably observed in CW-systems up to the investigated IR powers $\lesssim$ 30 W.

Although the qualitative discussion on the operation of a FIR laser and its dependence on the several parameters given above, is well agreed upon in the literature, the exact functional dependence of the parameters upon each other seems to depend on the level of sophistication used in the models. Moreover, experimental observations not seem to be unique either, which indicates that the performance of a FIR laser considerably depends upon the particular line chosen for investigation, or on the particular system used.

3 iv. Practical realization.

Attention is now turned to a description of the system, schematically outlined in Fig. 8. It is largely of the conventional type, home-built according to in-house design[16].

3 iv a. CO_2 laser.

The cavity of the CO_2 laser is of the hemispherical open resonator type, discussed in section 2 and the lasers practical performance relies to a large extent on passive cavity parameters. (In principle, an active cavity might differ from a passive one because the gain is nonuniform over the diameter. This causes a lensing effect, but it can be shown to be neglectable for practical purposes.)

The flat mirror consists of a reflective diffraction grating, having 150 lines/mm, to introduce a dispersive element in the cavity for line selection. The blaze angle is optimized to give ∿98 % reflection in first order at 10.6 µm and reflection varies only weakly over the CO_2 laser line branches. Its resolution is sufficient to obtain single line operation. Angle setting is obtained by a sensitive cantilevered micrometer drive with readout.

The spherical mirror consists of a ZnSe or CdTe material, with both sides coated and 10 m radius of curvature. This mirror acts as the outcoupling mirror and its transmissivity is determined by the coating. Coatings are available for all transmissivities desired over a broad range.

The active medium was contained inside a 13 mm ID, ∿1.5 m long discharge tube, ended by cylindrical cold cathodes and a common central anode. Each of the two discharge sections were series connected at the low end to a high voltage, high power pass tube and both circuits were parallel connected to a (stabilized) high voltage power supply. A glow discharge was maintained across both sections which were independently current controlled by controlling the grid voltage of the pass tubes. The sectional discharge reduces the operating high voltage and makes the discharge more homogeneous.

The tube is vacuumsealed by the output mirror on one side and a (flexibly mounted) ZnSe Brewster window (4) at the other. They are O-ring mounted and easily replaced. Both the outcoupling mirror and the grating are mounted in a two axis adjustable support for cavity alignment purposes. Both end mirror mountings are rigidly connected in a support frame consisting of three, 1 inch thick low thermal expansion super invar steel bars, on which the tube is also supported. These bars provide a passive stabilization of the cavity

The discharge tube is enclosed in a (tap)water cooled jacket. Without cooling, laser output drops by orders of magnitude after some time (∿1 hr) or even ceases at all, presumably because of thermal population of the upper and lower laser levels (which are only some 1000 K above groundstate)

The gasmixture, roughly consisting of 10% CO_2, 70% He and 20% N_2 (partial pressure composition) is continuously pumped through the tube at slow speed, mainly to keep constant operating conditions and relaxing vacuum requirements at the window vacuum seals (air may replace the function of pure N_2 gas). The gasmixture composition can be continuously controlled and optimized while it is taken from three separate flow controlled supply sources, containing the constituent gases (not shown in Fig. 8).

Typical operating conditions of the discharge are $\lesssim$ 10 kV, 20 mA at a total pressure of 20-25 Torr.

The distance between the mirrors is about 2 m, corresponding to a free spectral range for the cavity of $\sim$75 MHz (eq. 5). This FSR is much smaller than the bandwidth over which the CO_2 laser has gain ($\sim$width of the (broadened) molecular transition, $\sim$several 100 MHz). Because the transition is homogeneously broadened (CW), lasing occurs ideally on one longitudinal mode (the lasing mode quenching the gain at the other modes), the frequency of which is centered at the gain curve maximum. The frequency of the CO_2 laser therefore is tunable over the FSR $\sim$75 MHz. For this purpose, the back mirror (grating) is mounted on a piëzo-electric translator allowing easy cavity displacement over about three half-wavelengths.

Assuming the effective mirror radius to be given by the tube diameter, the Fresnel number for this cavity is $N\sim2$ at the laser wavelength. This Fresnel number together with the value of the parameter $|1-2d/R|\sim0.6$ (see section 2) lead to an expected cavity diffraction loss per pass^2 of $\lesssim0.1\%$ for the TEM_{00} mode. The (Gaussian) width of the beam diameter leaving the laser is found from (13a) as $2w_1\sim8$ mm. Because of the high gain of the CO_2 laser, output coupling is usually large compared to the other losses in the cavity (reflection losses at the grating and Brewster window, together a few percent per pass). Under this condition, output power is not very dependent on the transmissivity T of the outcoupling mirror (see (19), see also ref. 17). To obtain laser oscillation on as many transitions as possible in the high and low J regions of the P and R branches, where the gain is low, a relative high reflection ($\sim$80%) should be chosen for the outcoupling mirror. CO_2 laser output can be obtained in a single mode which is very close to the fundamental Gaussian TEM_{00}. Good mode quality however seems to put some stringent requirements on quality of the cavity components (output mirror, Brewster window, grating), cavity alignment, discharge conditions (distribution, dimension and stability of the glow area at the cathodes) etc., all conditions which are not straightforward to control. The system therefore rather easily tends to oscillate in other or impure modes. A ring shaped mode (with or without a central bright dot), the so called doughnut, with cavity distance dependent ring diameter, turns out to be a favorite alternative for the TEM_{00}.

It was felt that transverse mode control was easier accomplished when using a higher transmissivity output mirror (T∿40%) at the stronger lines. Because diffraction losses are strongly mode dependent, the additional increase of reflection losses is expected to give stronger mode discrimination in favour of the lowest loss mode (TEM_{00}). Still higher transmissivities (which are known to be acceptible for the strong lines as far as gain is concerned) may be even more favourable, but have not been tried. A larger radius of curvature for the output mirror, e.g. 20 m instead of 10 m (with a larger value of $|1-2d/R|$ and so increased diffraction losses), might have a similar effect in increasing the ease of mode control. It only slightly increases the beam diameter (9 mm instead of 8), while the TEM_{00} diffraction loss is still negligibly low ($\lesssim 0.1\%$). Typical output power of the CO_2 laser is in the order of 40 W in the fundamental mode.

3 iv b. FIR laser.

The cavity used for the FIR laser in the early work of Chang and Bridges[7] was of the Fabry Perot type as well. There are however considerable advantages in using a waveguide type resonator, introduced for the FIR laser by Hodges et al.[18] It consists of a hollow tube, either metallic or dielectric, with parallel flat mirrors at or near the ends. Typical dimensions and operating wavelengths, applying also for the laser presently described (Fig. 8), correspond to Fresnel numbers in the order of 0.2. This would correspond to free space diffraction losses in the order of 100% per pass, which clearly shows that waveguiding by the tubes is crucial for the operation of the laser. The transverse mode simply corresponds to the propagating waveguide modes, the end mirrors serving to obtain a standing wave pattern. Low propagation loss is obtained when the tube diameter 2a is much larger than the wavelength λ. Such highly overmoded waveguides are useless as passive waveguides because single mode operation is impossible to maintain. (It should rather be treated as a lightpipe than as a resonant waveguide). When part of an active laser cavity however, only the very lowest loss modes, corresponding to very low order modes, are selectively excited and sustained. Because of its large dimension ($\lambda \ll a$), the guide wavelength for the low order modes is very close to the wavelength in free space and longitudinal mode behaviour is controlled by the mirror spacing similar as for a free space Fabry Perot.

In the present experiments, either a 25 mm ID gold coated metal guide was used or a 13 mm ID quartz waveguide. The mode behaviour of a metal waveguide is well known from microwave theory, while the hollow dielectric waveguide modes have been first calculated by Marcatili and Schmeltzer[19]. There are some advantages of using a dielectric waveguide over a metal waveguide: higher losses at unwanted modes, lower (fundamental mode) coupling losses when the mirror is placed at some distance from the guide end, more favourable mode pattern for the lowest loss (EH_{11}) mode. A review of (dielectric) waveguide lasers can be found in ref. 20. The propagation losses for the dielectric guide scale as λ^2/a^3 and the 13 mm quartz guide can be used up to $\lambda \sim 200$ µm to have propagation losses $\lesssim 1\%$ per pass for the 1 m cavity[20] length used. The maximum guide diameter which could be used in the system of Fig. 8 was about 25 mm ID so that dielectric guides would be useful for $\lambda \lesssim 600$ µm. The lower loss metal guide therefore was chosen to use the laser at large wavelengths. Although guides could be relatively easy changed in the present set-up without re-alignment, this was often found inconvenient because an experimental arrangement behind the laser limited the free acces or because of time consuming pump down cycles. The metal guide therefore was often used also for short wavelengths.

The mirrors were gold coated and flat and adjustable for alignment. Simple hole output coupling from a 2.5 mm diameter hole in the back mirror was employed. Because in the present experiments the radiation was always further transported with lightpipes, output beam quality (deteriorated by hole coupling and possibly also by using a metal guide) was not of great importance. As far as beam quality is concerned, other types of output coupling[21] (mesh mirrors, hybrid couplers) might be more favourable, but are more complicated. Because the tuning range of a FIR laser line ($\sim$10 MHz) is much smaller than the cavity FSR ($\sim$150 MHz for the 1 m cavity length), one of the mirrors must be translatable to tune the cavity on a laserfrequency. This was accomplished by a large travel (2.5 mm) sensitive (50 µm/turn) differential screw, which could be either manually controlled or motor driven outside the vacuum enclosure. The vacuum housing of the FIR laser is closed with a quartz window at the exit, which transmits the FIR radiation and absorbs residual CO_2 radiation. A Brewster angled window (CdTe) closes the vacuum housing at the entrance side. The gas is continuously flown through the laser at very slow pumping speed, maintained by a rotary pump.

The CO_2 beam was directed into the FIR laser along the axis using an $\sim$f/100 focussing mirror through a 1 mm incoupling hole in the entrance mirror. At the low pressures required to obtain FIR laser action, the absorption coefficient for CO_2 radiation at (near) a resonant transition is general-

ly very low: 0.01 m^{-1}Torr^{-1} at best and for some gases or under saturation conditions even one or two orders of magnitude lower. Complete absorption of the CO_2 beam therefore would require a number of roundtrips for the CO_2 beam in excess of some 100. This number is by far not compatible with the cavity properties of the FIR cavity at the CO_2 laser wavelength in practice.
The CO_2 radiation suffers from reflection losses at the in- and outcoupling holes of the mirrors. The beam is focussed through the incoupling hole, causing it to diverge in the cavity with consequent beam walk out and absorption in the side walls. The actual and effective number of roundtrips is therefore rather in the order of a few. This makes α_{IR} in eq. (23) and (24) much larger than γL ($\gamma L/\alpha_{IR} < 10^{-1}$) and is the cause for the low efficiency of the FIR laser.

The scaling law (17) was found to be obeyed roughly for the two tube diameters used and was verified more systematically previously[13] (Accidentally, this is a rather general scaling law for waveguide gas lasers, also for other excitation mechanisms[20]). With this scaling law, $\gamma \propto p$ and $\gamma L \ll \alpha_{IR}$, one might expect a power output proportional to D^{-1}, but no significant diameter dependence was found in earlier work[13]. In the present system, at short wavelengths (<200 μm) a considerable increase in power output (factor 2 to 3) was observed with the 13 mm quartz tube as compared to the 25 mm metallic tube. A proper output power scaling with diameter however has to include the waveguide dependent losses, both at the IR and FIR wavelengths[13].

3 iv c. Microphone.

An extremely simple, but yet highly useful controlling aid is the microphone (13) shown in Fig. 8, which acts as a spectrophone when the pump beam is chopped, due to the photo-accoustic (also called opto-accoustic) effect. As the pump beam is absorbed, small temperature changes are induced in the gas with consequent pressure variations. These pressure variations are detected by the microphone.

The photo-accoustic effect, though already discovered in 1880 by Bell, has only after the advent of strong lasersources become exploited as a spectroscopic tool, but is by now one of the most sensitive techniques for doing spectroscopy on weakly absorbing gases or even opaque solids or liquids (in combination with a gas cell). It was first applied in conjunction with the OPFIR laser by Busse et al.[22]. A separate gas cell in the CO_2 beam path, filled with the FIR laser gas can be used[22] or it may be connected directly to the FIR cavity[23]. In the system of Fig. 8 it was mounted in a spare room of the vacuum housing of the FIR laser, well outside the FIR cavity and at considerable distance from the incoming CO_2 beam. A simple commercial electret type of microphone was by far sensitive enough to serve the purpose. The signals from the microphone were amplified and rectified and phase sensitive detection could be used optionally.

The location of the microphone was not optimized. In principle however some care is required because at the location used, a reasonable high background level was always present, due to nonresonant heating of the gas or indirect heating via tube walls, mirrors etc.. The heating due to resonant absorption by the gas was always sufficiently high (∿factor 1.5) above the background that the microphone was useful. For use at the very low pressure, a small hole had to be drilled in the microphone housing for pressure relief to prevent the active membrane from breaking. (This hole might reduce the sensitivity somewhat). Obviously, sensitivity strongly reduces for decreasing pressure (also for the background signal) but was fair enough at the lowest laser gaspressures used. In a separate gas cell the pressure might be independently chosen from the laser gas pressure, but is clearly more inconvenient. The chopping frequency for largest respons turned out to be about 100 Hz, but operation at much lower or higher chopping frequencies was possible. At the low chopping frequencies (∿10 Hz) which had to be used in some of the experiments, only the transients near the leading and trailing edges of the square waves were detected. At high frequencies, there was apparently a roll-off, possibly caused by thermal time constants of the system.

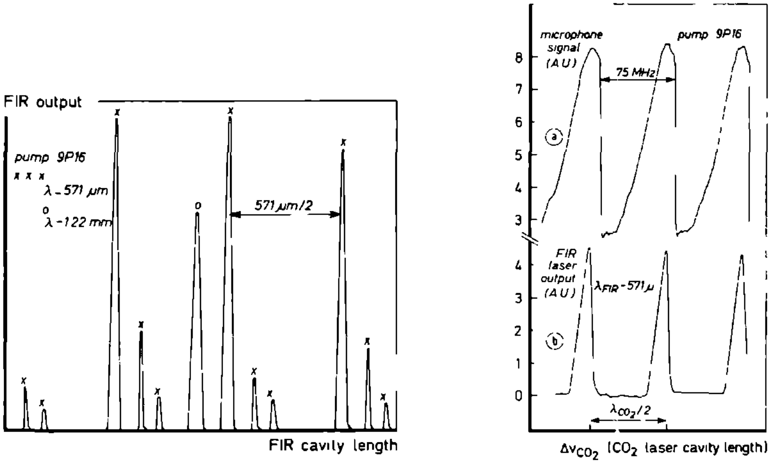
The usefulness of the microphone is clear. First of all, it is in no way interactive with the laser operation, neither in positive, nor in negative sense. It does not interrupt the CO_2 beam path and takes no extra power from it. For a given FIR gas, an absorbed CO_2 laser line is easily found from observing the microphone signal as a function of grating position. For the more familiar FIR laser lines and -gases, this is enough to identify the pump transition. Moreover, the CO_2 laser line absorption spectra of several of the more important lasergases have been published[24], so that the grating setting can be calibrated if wanted. The absorption frequency width in principle might be narrower than the CO_2 laser FSR, so that frequency tuning within a given line is required. For this purpose, the lasercavity length might be modulated during rotation of the grating[24], or one has to operate the laser in an (unstable) multi-(transverse)mode way[24], which also covers a broader range of frequencies. Simple rotation in a nominal single mode was found always sufficient however, presumably because the rotation automatically introduces a translation, so that the complete frequency range of a given laser line is scanned.

The tuning range of a (CW) FIR laser is much smaller than the FIR cavity FSR, so that FIR lasing occurs only for rather well defined settings of the FIR cavity mirrors. Also the CO_2 laser must generally be tuned within a frequency interval much narrower than its FSR to obtain FIR gain. One of the major advantages of the microphone monitor is that the CO_2 laser can be tuned independently of the FIR cavity tuning. Both the easy line selection (requiring no monochromator at all) and easy, independent CO_2 laser frequency tuning, facilitate routinuous use of the laser system and enable fast changes of FIR wavelengths. In addition, the microphone signal can be used in a feedback to stabilize the CO_2 laser at the absorption frequency[22].

3 v. Performance of the system.

Some typical performance data of the system are shown in Fig. 11. In Fig. 11a a FIR cavity scan is displayed when the CO_2 laser frequency and FIR gaspressure are optimized to give FIR output at some known lines of CH_3OH. The FIR power is detected by a pyro-electric detector. The peaks denoted by × are the dominant modes of the 571 μm FIR line. Presumably, but not verified, they correspond to TE_{11} modes of the metallic waveguide. The primed crosses correspond to some higher order low loss waveguide modes of the same frequency. The peak marked with o is a lasing mode of the λ=1.22 mm line as has been independently verified. The number of modes depends strongly on cavity parameters and on the specific line. It ranges from a few, as in Fig. 11a, to very many which may overlap each other (57, 47 μm lines CH_3OD, ∿400 μm lines HCOOH).

Both the 571 μm and 1.22 mm line are pumped by the same 9P16 CO_2 laser line and they are known to result from the same IR absorption transition[25]. Two other FIR lines are known to exist for 9P16 pumped CH_3OH[25], but they result from a different pump transition at a different frequency within the 9P16 tuning range and therefore they are not present in Fig. 11a, where apparently only one of the two transitions is pumped. It was not possible to have the cavity tuned simultaneously on both lines, but as they both result from the same upper level[25], they will be highly competetive and simultaneous lasing is not expected. Note that each mode has a rather narrow width and no output is observed between the modes, so the cavity has to be tuned rather



(a)

(b)

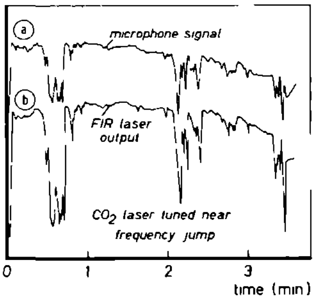


Fig. 11. (a): Typical FIR laser cavity scan with CO_2 laser tuned at optimum absorption. (b): Microphone signal (a) and FIR laser power output (b) as a function of CO_2 laser frequency (actually as a function of piezo-translator voltage, corresponding to laser mirror displacement). The FIR cavity is tuned at resonance for the FIR frequency.
(c): Comparison of microphone signal and FIR laser output for a free running laser system, with the CO_2 laser tuned near the end of its FSR.

(c)

accurately. This is a very characteristic feature of submillimetre wave lasers and is also the case for discharge pumped (∿HCN) lasers.

Fig. 11b shows the behaviour as a function of CO_2 laser frequency. For a given mode at resonance, the ratio d/λ is constant according to eqs. (1b) or (2), so that the cavity displacement is directly proportional to the frequency. The CO_2 laser has a collisional broadened bandwidth of ∿100 MHz at a pressure of ∿20 Torr[20], which is roughly equal to the gain- or oscillation bandwidth. (Actually, the tuning range may be either larger or smaller than the homogeneous molecular linewidth, depending on the level of laser saturation at the line center, see ref. 26).This bandwidth is not too much larger than the cavity FSR (75 MHz), so that the output power is not completely constant as a function of cavity tuning, but may vary by some tens of percent over the FSR. (this should be compared to the very narrow lines of the FIR laser, see Fig. 11a). This power variation is not shown in Fig. 11b. The upper curve of Fig. 11b gives the microphone signal and is proportional to the IR absorption by the FIR gas. The very assymmetric curves result from the fact that the maximum of the CO_2 gain curve does not coïncide with the absorption center frequency. As a result, absorption increases if the CO_2 laser is tuned further from line center and a maximum in absorption is found at a frequency near the end of the FSR. With further displacement of the cavity, the resonant oscillation jumps to the next longitudinal cavity mode, with corresponding frequency jump equal to FSR. The frequency jump gives rise to a nearly discontinuous drop in absorption. The discontinuous absorption drops correspond to the limits of the FSR, the CO_2 line center is about midway the FSR and the frequency offset with respect to CO_2 linecenter for optimum pumping can be read from Fig. 11b.

The lower curve gives the FIR output when the FIR cavity is tuned at resonance. It is seen that lasing occurs only very near the absorption maximum, over a limited range of the CO_2 laser spectral tuning range (FSR), within a given pumpline. The CO_2 laser frequency width, over which FIR lasing occurs, depends very much on the actual system used, as well as on the specific FIR line. Some FIR lines, which are nearly resonantly pumped such as the 118.8 μm line, may lase over the complete tuning range of the CO_2 laser, but the behaviour of Fig,11b is more representative for a general FIR line.

As expected, the FIR laser becomes very unstable when the CO_2 laser is tuned close to the frequency jump (see Fig. 11c). Accidental system instabilities then cause large CO_2 laser frequency instabilities with corresponding consequences for absorption, FIR gain and FIR output power. Note from Fig. 11c that the FIR output instability closely follows the absorption fluctuations represented by the microphone signal. Fig. 11c also illustrates the potential use of the microphone signal to (frequency)stabilize the CO_2 laser at optimum pumping, thereby stabilizing the FIR output.

Fig. 12 shows a FIR cavity scan at another FIR wavelength (λ=70.5 μm), together with the simultaneously recorded microphone signal. The microphone signal is rippled with ripple distance corresponding to $\lambda_{CO_2}/2$ and results

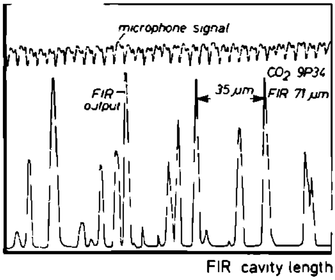


Fig. 12. Microphone signal and FIR laser output as a function of FIR laser cavity length. The CO_2 laser is tuned at optimal pumping. The periodic dips in the microphone signal correspond to IR resonances in the FIR laser cavity.

from cavity resonances of the circulating pump beam in the FIR cavity. Such a ripple is always observed, but the amplitude seems very much dependent on the actual cavity parameters and alignment conditions. It is typically between $\sim$5% and $\sim$20%. Note that the microphone signal does not represent the actual modulation of the circulating pump power in the FIR cavity, because it is mounted well outside the FIR cavity and considerable pressure wave impedance is expected between the cavity and the location of the microphone. It presumably only represents the reflected IR radiation from the FIR cavity which adds to the absorption contribution of the incoming beam. Note that the modulation appears as dips indeed, suggesting reflection minima. Occasionally, the modulation of the circulating pump power was measured directly by measuring the reflected IR power from the FIR cavity (not lasing and at atmospheric pressure) and was found to be in excess of 50%.

As discussed previously, a high cavity Finesse at the IR wavelength is desired for efficient operation (eqs. (23) and (24)), but does require that the cavity is tuned at resonance for the FIR wavelength as well as for the IR wavelength. Because the two wavelengths are well enough separated, this can be easily accomplished, either within a given tuning range of a FIR line (if corresponding to a distance larger than $\lambda_{CO_2}/2$) or by chosing a proper longitudinal FIR mode. (Note that a simultaneous resonance at two wavelengths λ_1 and λ_2 occurs periodically at distances separated by $\lambda_1\lambda_2/2(\lambda_1-\lambda_2)$). Using more sophisticated FIR mirrors ("hybrid couplers"), having high reflection at the IR wavelength, has indeed been found to strongly modulate the FIR output power due to IR longitudinal cavity resonances[21].

The ripple pattern of Fig. 12 may be of some use to determine the wavelength of the pump beam. Because many ripples can be counted, it was found that the wavelength determination is accurate enough ($\lesssim$ 1%) to identify the pump transition (of the CO_2 laser). Reversely, if the pump transition is known, the ripple pattern can be used to calibrate the FIR cavity displacement and FIR wavelengths. The latter application was previously found to be useful[27].

The tuning range of a FIR laser line can be most conveniently measured by observing a single mode such as in Fig. 11a on an expanded scale. Fig. 13 is an example for this at the longest wavelength which has been used in the present system. Distance is converted to frequency via $\Delta\nu=(\nu/d)\cdot\Delta d$ ((1b) ,(2))

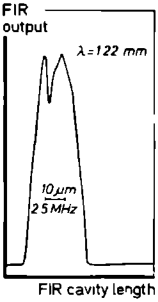


Fig. 13. FIR laser output as a function of cavity length for a single mode. The CO_2 *laser is tuned at optimal pumping. ("Tuning curve" of a FIR laser line).*

where Δd is the width on a length scale as experimentally observed, ν the center frequency and d the mirror distance which is known sufficiently well. Because at the FIR wavelengths , the molecular transition is predominantly collisional broadened and the oscillation bandwidth will be in the order of the molecular transition bandwidth[26], $\Delta\nu$ has no well defined frequency dependence. It is roughly constant, although it does depend on the parameters of the specific transition, cavity parameters, pressure etc. Therefore Δd is roughly inversely proportional to frequency. It was found that the tuning range of a FIR line at laserfrequencies ranging from 250-7000 GHz always lies in the range between some 5 and 15 MHz. These numbers are consistent with the molecular transition bandwidth, which is known to be in the order of 40 MHz/Torr[10]. The relative frequency stability $\Delta\nu/\nu$ of a lasing line for a fixed cavity setting is equal to the cavity distance stability $\Delta d/d$ (If refractive index variations are neglected). It is not directly determined, but estimated to be $<10^{-7}$ (presumably smaller for the CO_2 laser which has a better passively stabilized cavity), which corresponds to $\lesssim$few MHz for the CO_2 laser and <100 kHz for the FIR laser (both unstabilized). The ultimate natural laser linewidth for the ideal cavity is very small $<10^{-3}$ Hz as can be estimated from passive Q and extracted power level. Frequency stability of the FIR laser is of no importance at all for the experiments described in this work. The dip present in the scan of Fig. 13, which is usually present, is presumably caused by the standing wave pattern of the pump beam. Such dips sometimes were also present at other wavelengths (notably 70.5 μm) and might here occasionally go to zero. For short wavelengths the FIR tuning range is too small to allow a standing wave IR modulation of the FIR output at each longitudinal mode (see e.g. Fig. 12, where the dip however is not present), while each mode is apparently splitted by the dip. A possible cause for the dip at short wavelengths will be discussed subsequently.

3 vi. Effects influencing the FIR laser stability.

When using the FIR laser as a source of radiation for other experiments, as in this thesis, only two parameters of the system are important: the power output and its stability. Power output has been discussed in the preceeding and practical values were (depending on transiton) in the range 1-50 mW (CW), which was conveniently high. For the experiments, the laser was nearly always used in a free running mode without stabilization. Options for active stabilization, e.g. using the microphone output were available, but it was found that under the usual optimal laser conditions not much was gained with additional stabilization, while an uncontrolled unstable laser was difficult (or impossible) to stabilize. Long term modulation or drift problems were more easily resolved by a normalizing technique with respect to a reference beam (see chapter III). Typical short term amplitude stability (∿min) is in the order of a few percent (1s integration time). This stability is at least comparable or better than the best published stability behaviour of comparable (stabilized) systems[11,13,22].

The (power)stability of a FIR laser is completely dominated in practice by the frequency stability of the CO_2 laser. A great number of factors govern the CO_2 laser frequency dependence of the power, as well as the CO_2 laser frequency itself. A situation may exist in which many transverse modes of the CO_2 laser have low losses such that they will easily oscillate. For the cavity configuration R=10 m, d=2 m, about 7 modes with the same axial mode number exist having frequencies equally divided over a free spectral range (eq. (15)). Higher order modes of different axial mode number, oscillating within the same frequency range, have in general frequencies incommensurable with the first set of transverse modes. Thus for a given cavity setting, the modes can cover an almost continuous frequency range (in principle in the order of the total oscillation bandwidth of the CO_2 laser). Because the lasing transition is homogeneously broadened, the modes are strongly coupled and multimode operation is not likely. If for some reason, there is insufficient mode loss/gain discrimination, the laser frequency may become unstable and difficult to control or stabilize. The frequency behaviour is arbitrarily controlled by accidental conditions: cavity stability, discharge stability, coupling of radiation field and discharge etc. Although such multimode behaviour is useful for some applications[24], it is detrimental for obtaining a useful FIR power output.

A natural mode discrimination is given by the frequency dependent gain (although counteracted by saturation effects for high laser intensities[26]) and by the mode dependent diffraction losses. Single mode operation however still puts some demands on cavity design, stability, discharge conditions, quality of optical components, alignment etc., which conditions have been met for the present system[28]. Single, tunable mode operation (see e.g. Fig. 11b) was always employed to obtain stable FIR output. This mode is preferably the TEM_{00} (which is the most natural mode for a highly stable system) which has the most ideal intensity distribution. In practice, satisfactory FIR laser performance could also be obtained either in another mode or from an apparently somewhat distorted Gaussian (presumably due to contamination of optical components). Evidently, frequency stabilization of the CO_2 laser[13] is possible only with single mode operation.

The pumpfrequency dependence of the FIR power is obviously severely dependent upon the amount of off-resonance pumping. If $\delta\nu_{IR}$ is the frequency difference between the CO_2 frequency and the absorption line center, the absorption coefficient as a function of $\delta\nu_{IR}$ follows the Doppler profile from the Maxwellian velocity distribution[11]:

$$\gamma_0(\delta\nu_{IR}) = \gamma_0(0)\ \exp(-\delta\nu_{IR}^{\ 2}/4\delta\nu_D^{\ 2}) \tag{26}$$

where $\delta\nu_D^{\ 2}$ is the Doppler broadening at the pumpfrequency. The absorption, and so FIR performance and stability, sensitively depends upon $\delta\nu_{IR}$ and the stability of it.

It has been frequently reported[11,13] and shown[11], that reflection of IR radiation from the FIR cavity back into the CO_2 laser causes a frequency jitter of the CO_2 laser (and power fluctuations). No direct evidence for this effect was found (or looked for) in this work, but the analogous effect has been observed for the FIR laser as an annoying feature when testing interferometers closely behind the laser. If some of the FIR power coupled out was reflected back into the laser, it resulted in laser output power modulation as large as 20-30 % as a function of (unknown amount) power reflected back.

It is well known that under conditions of high pump intensity (saturation behaviour), a selected group of molecules is preferably excited, whose Doppler shifted frequency corresponds to the pump frequency ("hole burning" in the pumped transition, schematically indicated in the inset of Fig. 10). At low intensity, the excitation is rapidly shared by all molecules because of velocity relaxation. Evidence for hole burning in the absorption transition came already from a theoretical analysis of pulsed OPFIR laser experiments[29]. It has since been intensively investigated, theoretically[15,30-34] as well as experimentally[11,31,33,35-38] and particularly for CW systems. Hole burning effectively reduces the number of participating molecules to the FIR lasing process[29] and will make the FIR output even more CO_2-frequency dependent. These effects are especially important when the pump beam is circulating in the FIR cavity, as it should do for high efficiency ((23),(24)) and actually does in practical systems (note the standing wave pattern of the present system, evidenced by e.g. Fig.12, see also ref. 21). In this case, two holes are burned symmetrically displaced at both sides from the absorption center frequency by the frequency offset $\delta\nu_{IR}$. This can be nicely demonstrated for FIR lines, where the center frequency of the pump transition is accidentally within the tuning range of the CO_2 laser. As a function of CO_2 laser frequency, (an experiment corresponding to Fig. 11b) the FIR power output shows a dip at the center, corresponding to the collapse of the two symmetrically around center displaced frequency (velocity) groups into one central group. This dip was predicted[32] and observed[36,38] for several lines, amongst which the strong 118.8 µm line[36], which can be resonantly pumped. Because of its close analogy with the Lamb dip in inhomogeneously broadened gaslasers, the effect is called the transferred Lamb dip[36]. Note that neither the CO_2- nor the FIR laser are expected to have a Lamb dip of their own, as both are homogeneously broadened (predominantly). The dip was also directly observed in the absorption from the photo-accoustic signal[36].

An interesting case occurs at low FIR gaspressures and/or high FIR laser frequencies. The Doppler broadening may now be comparable to or exceed the collision broadening even at the FIR frequency. The laser transition becomes then somewhat inhomogeneously broadened. If the pump frequency has an offset $\delta\nu_{IR}$ from absorption line center, the "hole burning" in the pump transition is transferred as a spike (of ~collisional linewidth) in the FIR gain profile. This will cause the FIR emission to be amplified for propagation in one direction only. If the pump beam is travelling back and forth (which it normally does), two spikes are induced in the gain profile and the FIR emission is amplified in both directions, but by molecules of opposite velocity classes. Thus the FIR tuning curve shows a dip in the center, or is completely split in two peaks, the splitting being proportional to the pump frequency offset. This effect can be easily observed. It seems especially that the 70.5 μm line (CH_3OH) is apt to show this effect at frequency offsets only some 10 MHz where it has been observed[31,37].

The effect follows from rate equation models, taking hole burning into account as well as an inhomogeneously broadened FIR transition [15], but similar effects are also expected from quantummechanical two-photon Raman-type processes[31,33]. The anisotropic gain (unidirectional amplification) was also directly demonstrated in an unidirectionally pumped amplifier cell[37]. The splitting of the FIR tuning curve was also demonstrated for the 118.8 μm CH_3OH line[11].

It is believed that this effect also accounts for the apparent splitting of the 70.5 μm line tuning curve which was sometimes observed with the present system. Although it seems possible to near resonantly pump this transition[31,37] a small (~10 MHz) accidental pump offset (due to an improperly tuned or drifted CO_2 laser) is not unreasonable. It is also of interest whether these effects are responsible for the marked different behaviour of the 70.5 and 118.8 μm lines which was found especially when using the metal waveguide. Both lines are among the best known strongest FIR lines and have relatively strong pump absorption. Their wavelengths are comparable and they are pumped by adjacent CO_2 transitions (9P34 and 9P36 respectively). Yet, the 70.5 μm line appeared to be much more sensitive to external conditions than the 118.8. A severely misaligned FIR cavity, combined with an off-axis injected input beam, only weakly affected the 118.8 transition, but prevented the 70.5 line to lase at all. Under the same condition, the 699.5 μm CH_3OH line, which results from the same pump transition and upper laser level[25] as the 70.5 line, was easily found to lase, while the latter line is weaker than the 70.5 line for aligned conditions.

A difference between the 70.5 line and the other two mentioned lines is that the 118.8 and 699.5 lines are clearly not expected to suffer much from gain anisotropy: 118.8[11] is known to be resonantly pumped, operates at a somewhat higher pressure and is not extremely sensitive to pump frequency;

the 699.5 line is certainly not Doppler broadened. Possibly, some spatial separation in forward and backward pump intensity, with associated spatially different regions of gain for the forward and backward FIR intensity might also affect the operation and stability of a FIR laser.

It may be concluded from the foregoing discussion that effects of inhomogeneous broadening of the FIR transition adversely affect the stability performance. These effects are thus more important at higher frequencies (larger FIR Doppler broadening)[11]. Although the present system was not intended to be used at extremely short wavelengths (and optical components such as the exit quartz window are not adapted for it), it was occasionally used at a wavelength as short as 41.7 μm (see chapter III). Stable operation however was rather difficult to obtain at this wavelength.

To decrease effects of inhomogeneous line broadening, the laser should be operated at the highest possible pressures, i.e. in the smallest waveguide compatible with the propagation losses at that wavelength. The guide diameter is not expected to influence saturation effects very much, as both intensity and pressure are inversely linear with diameter.

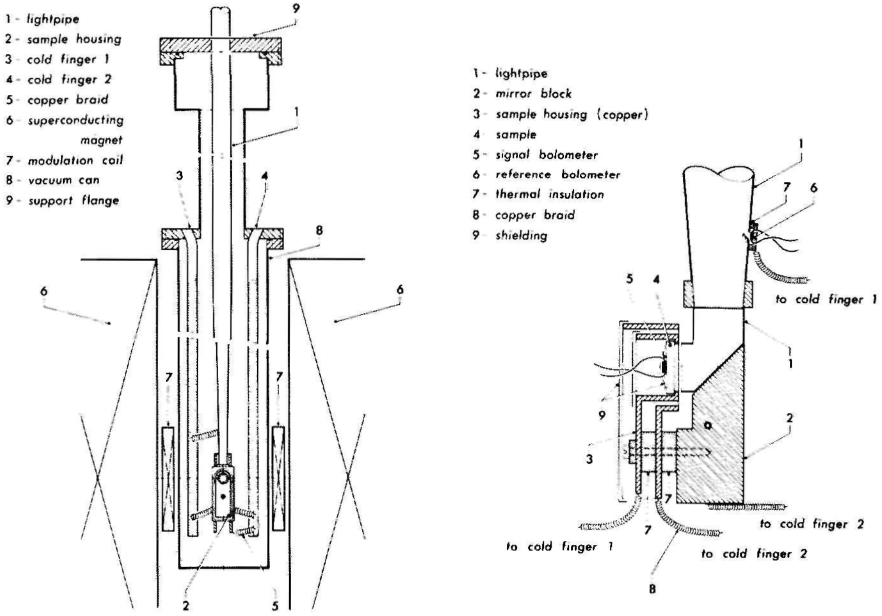
4. Detection of radiation absorption.

The most straightforward way to obtain spectroscopic information on opaque samples is to measure the reflectivity. For highly reflecting samples (e.g. metals or high refractive index materials), the reflectivity is not very sensitive to small changes in the material properties as the corresponding changes in reflectivity have to be detected against a high reflection background. Additional techniques therefore are required to increase the sensitivity such as the use of cavities and/or modulation methods. The most sensitive way under these circumstances is to measure the absorption directly. A possible way would be the use of the photo-accoustic effect, discussed in the previous section. In the present work, we have however adopted a more familiar method, which consists of calorimetric detection.

4 i. Calorimetric method.

An outline of the experimental set up is shown in Fig. 14. The absorption has to be detected as a function of magnetic field. The magnetic fields were generated by a superconducting coil which was supported in the same liquid helium bath as the cryogenic insert (see Fig. 14a). To facilitate the possibility for controlled heat leaks and thermal insulation, the sample mounting and support was enclosed in a vacuum can. Two helium filled cold fingers subtended in the vacuum can to provide the cooling capability. Cooling of the different parts of the assembly was obtained by connecting them to the cold fingers, using copper braid. The braid is tightly attached to the parts over some length with GE 7031 varnish to provide the desired cooling capability.

A detail of the sample mounting is shown in Fig. 14b. The radiation enters the lower part of the insert via an oversized lightpipe (9 mm ID) ending in a conical transit element to reduce the final beam size to some



(a) *(b)*

Fig. 14. (a): Lower part of the low temperature absorption measurement system.
(b): Sample housing and sample mounting. (Detail of (a)).

6 mm. Because the sample surface was parallel to the magnetic field (Voigt configuration), a 90° bending piece was used just in front of the sample. The sample mounting was thermally insulated from the support frame via a foam material spacer (7) (Rohacel) and a synthetic material mounting screw (Delrin). A foam material clearly has the highest thermal insulating properties over other synthetic materials, while Rohacel is capable of providing at the same time some required mechanical rigidness. It is also conveniently mechanically processable (e.g. machining). A bolometer (thermometer) was glued at the back of the sample to register the temperature (variations) of the sample.

4 ii. Bolometer.

Because the bolometer is in a magnetic field which is continuously varied in the experiment, the bolometer (it should more properly called a thermometer in the present application) must not be very magnetic field dependent. This requirement makes the highly sensitive doped Ge or Si bolometers less suited for the present use, but he carbon type resistor is well known to be applicable here. A carbon resistor depends for its low temperature behaviour strongly upon the composition in a way analogous to a normal semiconductor. Their useful properties (high temperature dependence of the resistance at low temperature) is not intentionally controlled but the standard carbon resistor as manufactured (previously) by the Allen&Bradley company (A&B) are widely known for their strong temperature dependence at low temperature.

A 220 Ω, 1/8 W A&B resistor was used as the bolometer in the present experiments. For this purpose, the paint was scratched off down to the carbon body. One side was flattened and somewhat polished to obtain a large thermal contact area. Finally, it was somewhat thinned down to some 0.3 mm thickness and the resistance was about 1.5 times the original value. The original soldering electrodes were used, which were however cut off at $\sim$2 mm distance from the resistor. A mercury battery was connected to the bolometer in series with a cooled, wire wound resistor, with a resistance much larger than the bolometers resistance to provide a current bias. The resistance (changes) are thus simply proportional to the dc voltage across the bolometer. Thin manganin electrical wiring was used to the bolometer to reduce heat leak via the wires. Radiation induced changes in the bolometer voltage were phase sensitively detected at the frequency of the chopper, which modulates the CO_2

laser power. It is important to chop the CO_2 laser and not the FIR output beam, because in the latter situation, with the chopper close to the entrance lightpipe, a synchronous modulation of the background radiation reaching the bolometer occurs. A modulated CO_2 beam is moreover useful for the microphone (see section 3).

Typical bolometer operating conditions at the operating temperature (which varied somewhat from run to run) are $R \sim 150$ kΩ and dc bias ~ 2.7 μA. The operating temperature depends on the bath temperature and on the cooling of the sample against the residual heat leak (mainly radiation). The helium bath was pumped down at modest pumping capacity.

Two different magnets and dewars were used. Low field experiments were preferably done with a small 4T solenoid in a simple glass dewar, while measurements at higher fields were done with a 7T solenoid in a metal dewar. In the first system the lowest temperature was about 1.5 K and somewhat higher in the latter (1.9 K).

A rough guide for the sensitivity of an A&B resistor in this temperature region is given by:

$$S \equiv \frac{1}{R}\frac{dR}{dT} \simeq -(2\pm 1)\ K^{-1} \tag{27}$$

This sensitivity was not calibrated, but is taken from general literature values. The number (27) is reliable for rough estimates, but actual values of course depend on the specific resistor and its actual fabrication process, which is subject to changes. The sensitivity also increases with decreasing temperature.

4 iii. Sensitivity.

The sensitivity of a calorimetric radiation absorption technique relies on the use of phase sensitive detection (PSD). The thermal time constant of the system should therefore be compatible with a modulation frequency, which is useful for PSD ($\gtrsim$ Hz range).

A thermal analysis of the system is done, based on the heat continuity equation: applied heat input is either stored in the components having a finite heat capacity, or carried away to a heat sink (bath). The temperatures of the components can then be found for a given heat input. The heat continuity equations are obviously analogous to the current continuity equations for electrical circuits containing lumped resistors and capacitors and for the proper equivalent electrical circuit the solutions can be written down immediately, using the complex impedance analysis.

A suitable electrical circuit to analyze the system of Fig. 14b is given by Fig. 15. The heat power supplied to the sample from the radiation absorption at the (chopper)frequency ω may be stored by its heat capacity C. The sample is connected to the heat sink by a thermal resistance R. A bolometer with heat capacity C' is thermally connected to the sample by a thermal resistance R'. The bolometer is assumed to have no additional thermal contact with the bath, e.g. the heat conduction by the electrical leads is ignored. Obviously, such equivalent circuits as Fig. 15 can be made as fancy as desired, but as a lumped circuit is a simplification by itself, complex circuits should be considered with care. The circuit of Fig. 15 is realistic if thermal gradients across the sample and bolometer can be ignored.

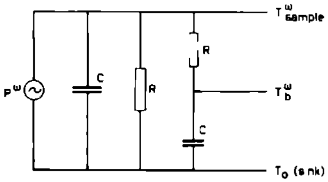


Fig. 15. Equivalent electrical circuit for a thermal analysis of the experimental configuration of Fig. 14(b).

The temperature of the bolometer T_b^{ω} (amplitude at frequency ω) is then given by:

$$|T_b^{\omega}/P^{\omega}| = |R/((1+j\omega RC)(1+j\omega R'C') + j\omega C'R)| \tag{28}$$

From (28) (or directly from Fig. 15) the maximum bolometer temperature (i.e. the sample temperature) is obtained for $\omega R'C' < 1$ (i.e. no temperature drop across R' compared to the temperature drop across C'). The bolometer is glued against the sample using GE varnish with an additional thin paper foil (cigarette paper) for electrical insulation.

An estimate for R' for such a thermal contact (neglecting the (presumably varnish impregnated) paper foil) may be obtained from a semi-empirical expression[39]:

$$R' = 2r_B/A + t/KA \tag{29}$$

with t thickness of the bond layer, K speçific thermal conductivity of the bonding material, A bonding area and r_B the contact resistance. For sufficiently thin bonding layers, the bulk resistivity of the bonding agent is ignorable. For a bond between copper surfaces r_B is found to be[39]:

$$r_B \simeq 7.5\ T^{-3}\ \mathrm{cm}^2\mathrm{K/W} \tag{30}$$

Eq. (30) is nearly independent of the bond material used (e.g. GE 7031 varnish, Apiezon N grease and other typical cryogenic bonding agents). The thermal conductivity K of such materials is extremely low:

$$K \sim c\ T^2 \tag{31}$$

with typical values for c: 10^{-5} W/cmK^3 for Apiezon N grease, or $\sim 3.5\ 10^{-5}$ W/cmK^3 for GE varnish,while the latter value is especially sensitive to preparation. Assuming (29) to be valid for contacts between other materials as well (e.g. Bi and graphite in the present case) it is found that the thermal resistance is dominated by the bulk resistance of the bonding material for thicknesses t in excess of ~ 3.5 µm at $T \sim 1.5$ K.Because the practical value of the thickness is not known, but is certainly not much smaller than 3.5 µm (e.g. due to the foil), the thermal resistance is difficult to estimate. With a specific heat capacity $\sim 10^{-4}$ J/cm^{3}K for graphite, the heat capacity for the $\sim 3\times 1\times 0.3$ mm^3 bolometer is obtained as $C' \sim 10^{-7}$ J/K. With this value, the condition $\omega R'C' \ll 1$ is easily fulfilled, even for rather thick varnish layers, at the typical low modulation frequency of ~ 10 Hz. In fact, the bolometer respons will be limited by its own intrinsic time constant to obtain a uniform temperature.

For estimates, it is useful to compare the wavelength of a temperature wave in a solid material with the dimensions of the sample. The wavelength is given by[40]:

$$\lambda_T = \sqrt{(8\pi^2 K_m)/(C_m \omega)} \tag{32}$$

where K_m and C_m are the specific heat conductivity and heat capacity respectively. With a typical very low conductivity of graphite $K_m \sim 10^{-4}$ W/cmK at $\lesssim 2$K, $\lambda_T \sim 1$ cm at the chopping frequency of 11 Hz. The intrinsic time constant will be of the order of the time required for a thermal wave to transit the sample and thus given by $d/v \overset{\sim}{=} d/\lambda_T \nu \sim 3$ ms, where d is a transverse thickness (perpendicular to the contacting area) of the bolometer. This time constant is still much lower than the inverse chopper frequency.

A similar time constant applies to the sample itself as well. Taking Bi as the example for a sample material, with reasonable values $C_m = 2\cdot 10^{-4}$ J/cm^{3}K and $K_m \overset{\sim}{=} 4$W/cmK (applying to reasonable pure Bi at ~ 1.5K), the wavelength is found to be $\lambda_T \sim 150$ cm at 11 Hz. With typical sample thickness 1.5 - 2 mm, the intrinsic time constant is $\sim 10^{-4}$ s, which justifies the assuption of a uniform temperature distribution across the sample on a time scale of the inverse chopper frequency. With the heat capacity of the typical Bi-sample (thickness $\sim$1.5 mm, diameter 8 mm), $C \sim 1.5 \cdot 10^{-5}$ J/K, much larger than the bolometer heat capacity, (28) reduces to

$$T_b^{\omega}/P^{\omega} = R/\sqrt{1+(\omega RC)^2} \tag{33}$$

The bolometer and its mounting behave thus rather ideal in that it follows the sample temperature unattenuated and does not limit the thermal time constant. According to the estimates, this is true within a large margin and depends not critically on the crude estimates for the values of the parameters.

It is now left to select and realize a proper value for R in (33). At a given frequency, the signal (T_b^{ω}) increases linearly with R at low R ($R \lesssim 1/\omega C$) and approaches a constant value ($1/\omega C$) for large R. A cut off frequency is defined as $\omega_c = 1/RC$. Although the signal does not depend on R above cut off, operation above cut off must be avoided as it causes the sample temperature to drift away when radiation is applied. The optimum value for R is thus the value just below cut off: $R \lesssim 1/\omega C$ The chopping frequency must be chosen to be useful for lock-in detection i.e. $\omega/2\pi \gtrsim 1$ Hz. A value of 11 Hz was chosen in practice. For reasons to be discussed, the sample was glued with its edges against the rim of a massive copper housing which was intentionally kept at the bath temperature and may be considered as the heat sink for the sample. The resistance R is thus formed by the bonding layer. Estimates for R may be based again on (29), but it turns out that the bulk term is not unimportant. The thickness of the bonding layer and the thermal conductivity of the bonding material are not known and the bonding area only very inaccurately, so that estimates are not very reliable. This point has therefore been tested in practice.

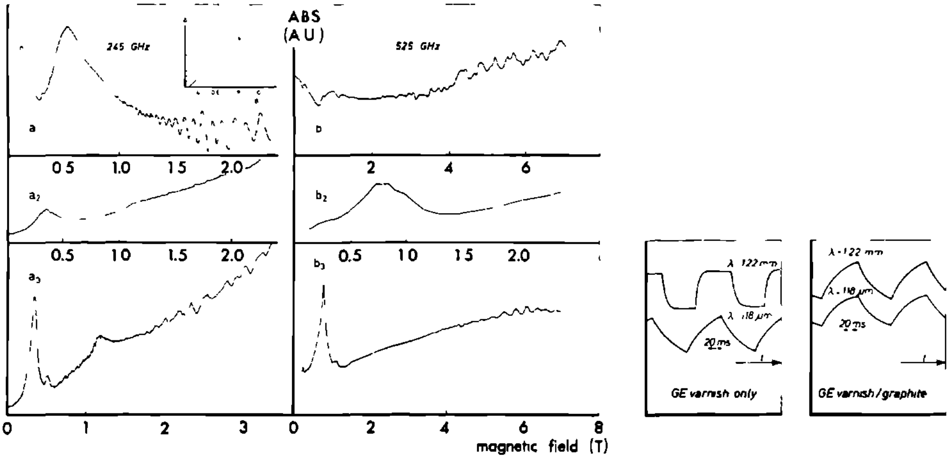
4 iv. Test measurements.

In early test measurements, it already turned out that a background signal, not due to direct sample heating, was generally present at the signal detector. Two obvious causes of this background are heatleaks from different parts of the assembly, which are heated also by the radiation, to the sample and stray radiation impinging directly to the bolometer. Heat leaks are rather simple to reduce. The sample housing is nearly perfectly isolated from the support frame by the foam material spacers. This very efficient isolation (power dissipation at the support frame, close to the sample housing had no effect on the sample temperature up to powers $\lesssim 25$ mW; at higher power dissipation, the sample temperature started to increase very slowly and is presumably caused by the slight heating of the He-bath itself) also relaxes the temperature requirements for the support frame and so facilitates cryogenic design. To reduce heating of the sample housing ((3) in Fig. 14b) and preventing stray radiation from heating the bolometer, the sample housing was completely shielded by an additional screen, enclosing the sample housing. Thermal cross talk has been further reduced by reserving one of the two cold fingers for the sample (and reference detector) solely. In this way, a simple and effective screening is obtained, which is better than required.

More severe and more difficult to avoid is the radiation leaking between the sample front face and its attachment to the holder (see Fig. 14b). Dummy test samples were found to be useful in tracing and measuring the leakage signals. High purity, but polycrystalline Bi plates serve very well as test samples. Bi is known to have some strong, easily detectable, absorption signals as a function of magnetic field. Even if these signals cannot be completely identified and are not very reproducible because of the polycrystalline nature of the samples, they are very suited for test purposes.

Fig. 16a presents some signals observed at the signal bolometer as a function of magnetic field at two laser frequencies and three different sample mountings. In all cases, as in the final experiments described in chapter III, the signal was normalized to a reference signal obtained from a second detector ((6) in Fig. 14b). For trace a_1 and b_1, the sample was glued against the holder using GE varnish only, in a_3 and b_3, the (dummy) sample was soldered (using 50° C melting point solder) to it and for the middle traces, a_2 and b_2, it was again glued to the housing with varnish, but now the varnish was mixed with finely grinded graphite powder. For the three mountings, the sample was independently positioned and mounted, so that the shape and positions of the signals should not be compared, only qualitative features are relevant. Note also the different field ranges displayed, as the data are taken from different series of measurements.

The point to be noted is that the peaks and high field variations in curves a_1 and b_1 are reversed in sign as compared to both the curves of



(a)

(b)

Fig. 16. (a): Typical detector signals as a function of magnetic field for a polycrystalline Bi-sample at two laserfrequencies 245 GHz (a) and 525 GHz (b). In a_1 and b_1 the sample was glued to the holder with GE varnish only, in a_2 and b_2 with GE varnish mixed with graphite powder and in a_3 and b_3 it was soldered to it. (b): Comparison of thermal time constants at high and low laserfrequencies for two different sample attachments.

a_2 and b_2 and a_3 and b_3. The lower two curves (indexed 2 and 3) display the behaviour which is expected for the absorption (see chapter III). The upper curves (indexed 1) can be explained assuming that the radiation leaks between the sample and the holder and the leakage signal reaching the bolometer completely dominates the heating signal. In leaking along the sample however, the radiation is attenuated by the (field dependent) sample absorption and so the absorption signal is present in the leakage signal with inverse sign. Soldering the sample against the holder is most straightforward to reduce leakage to zero. The varnish/graphite mixture evidently also provides sufficient attenuation to the radiation to effectively reduce the leakage. Only two low laserfrequencies are shown in Fig. 16a, but it was observed that leakage was never noticeable at high frequencies (short wavelengths << 0.5 mm), that is, for high frequencies the signals always had the "correct" absorption sign, independent of the bonding material, even for unpolluted varnish.

Fig. 16b illustrates an independent way in support of the previous observations. The ac bolometer signal due to the (chopped) radiation absorption is displayed on an oscilloscope, so that the respons times can be directly checked. When the sample was glued with GE varnish only, at long wavelengths (1.22 mm , 245 GHz) the relaxation time was much shorter than at short wavelengths. This means that at long wavelength the respons time is determined by the bolometer only (because the signal results from leakage) and by the sample at short wavelengths (because the signal results from sample heating). When leakage is prevented (by the varnish/graphite mixture), the relaxation time is short, both at long and short wavelengths.

An interesting feature are the high field oscillations present in most curves of Fig. 16a, but most nicely in (a_1). At high field the dielectric constant of the Bi sample becomes real and wave propagation in the sample occurs. For a compensated plasma such as Bi, they are of the Alfvèn type. Their wavelength (or dielectric constant) varies with magnetic field in a well known way. There are two classical manners to study the dispersion of the waves. Either the sample is thin enough and plane parallel, so that a standing wave pattern is generated across the sample thickness. Measurements of transmission, reflection or absorption then show a Fabry Perot modulation pattern as the magnetic field (wavelength of the waves) is varied. Another way is to have some transmission through the sample and at the same time some leakage signal around the sample (cohaerent radiation). When both the transmitted and leakage signal combine behind the sample, the total intensity is modulated because of the modulation of the phase of the transmitted wave, compared to the original (leakage) wave. This type of interference is called Rayleigh interference and the period differs a factor of two with the the Fabry-Perot period.

It is interesting to note that in previous investigations of Rayleigh interference in Bi at 0.5 mm wavelength[41] the low field structure had the same sign as in Fig. 16a, (a_1) and (b_1). For the oscillation in Fig. 16a, (a_1), it is not known whether they are Fabry Perot or Rayleigh interferences. Wave interference was generally seen in the dummy samples, also in cases where leakage was prevented (see e.g. curves (a_3) and (b_3)). The oscillations in (a_1) clearly belong to a well defined dispersion relation as indicated by the inset. Subsequent peak positions follow a 1/B behaviour, as expected for Alfvèn waves.

The oscillation pattern looks more generally however as in Fig. 17. The irregular behaviour is not noise but corresponds presumably to different periods arizing from different crystallites. For curve (a) of Fig.17 the sample was soldered and in curve (b) it was glued with GE varnish. Both curves are similar, having not the scaling behaviour with thickness

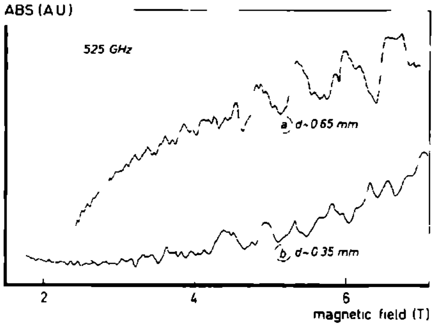


Fig. 17. Detector signal as a function of magnetic field for two polycrystalline Bi-samples. The oscillatory structure is a result of Alfvèn wave propagation.

expected for a single crystal series. The results for a soldered sample must be due to Fabry Perot resonances. It might be possible that the nice oscillations in (a_1) of Fig. 16a result from leakage, which predominantly occurs over a limited part of the rim of the sample, covering accidentally only one crystallite. Crystallites magnitude were observed by the eye to be in the order of a mm.

It is expected that the magnitude of the leakage depends strongly on the width and thickness of the varnish layer. The thickness is certainly very small compared to a wavelength at all wavelengths, so that a strong attenuation is expected. In fact, a parallel plate transmission line (strip line) of plate separation much smaller than a wavelength is sometimes used in sensitive absorption measurements in the far infrared[42] (although typical distances are larger than the present thickness of the varnish layer). This is believed to be the reason that the absorption features are so strongly present in the leakage signal. The total attenuation will further increase with varnish layer width. The width was chosen rather small (a few tenths of a mm, but may vary somewhat around the circumference of the sample), as this was expected to optimize the signal strength (heating) via optimized cooling. It is therefore comparable to, or even smaller than, a wavelength at long wavelengths. This explains that leakage was only severe at long wavelengths. The stronger diffraction effects at long wavelength will also add to a higher leakage level.

Covering a large part of the rim of the sample by the sample housing (using large samples) and/or exposing only a small area to the incoming radiation is impractical, respectively unwanted as it decreases the signal strength. Soldering of the sample is clearly undesired for actual measurement samples as it involves too crude a treatment which may adversely affect the quality of the sample. Moreover, it was found that the Bi sample dissolved somewhat in the (Bi rich alloy) low temperature solder.

For the final experiments (chapter III), the sample was always glued using the varnish/graphite mixture. This turned out to work satisfactory for the work described in chapter III. In general however, each glueïng is undesired as it causes strains in the sample upon thermal cycling. This problem is not easily circumvented. An alternative might be to provide the sample housing rim against which the sample is placed with a radiation absorbing material and to press the sample against the rim without sticking to it. A separating mylar foil might serve to prevent the sample from sticking. This is essentially the method employed by Blewitt et al.[43] in transmission measurements. They mentioned however that leakage was still insufficiently decreased at very low (transmission) signal strengths. In a calorimetric detection system however, another method would then have to be found to cool the sample. A firm contact with the sample is always unavoidable, but the attachment of a cooling rod at the back of the sample (where the bolometer already is mounted) might be preferable over a contact at the front face rim.

Cooling with a rod in principle might give a better heat removal control by adjusting the length and/or thickness of the wire (via $1/R=K_m A/l$ with K_m specific heat conductivity of the wire material, A crossectional area and l either the length between the sample and sink or $\sim$ the wavelength λ_T (32), which one is lower). In practical tests however, the contact resistance between wire and sample was considerable and a rather bulky wire was required to obtain a suitable thermal time constant.

It should be noted that eq. (33) was obtained from a heat balance equation for the sample only, irrespective of the actual realization of the thermal resistance R. The heat capacity of the cooling wire therefore does not directly influence the thermal relaxation time of the sample+bolometer system, even if the contact resistance were small.

An easy way of cooling would be to work in a He gas atmosphere, or even in liquid helium (above or below the λ point). (In the latter case, no control of heat resistance is possible). This way of cooling was not tested, but it is expected to be larger than optimum (see also ref. 5). Gas cooling is proportional to pressure (at low pressures, explicit formulas see e.g. ref. 44). Gas cooling was incidentally tried, but was found to give less thermal isolation between the sample and the assembly. It might however be a reasonable alternative (with the sample in a proper enclosure to reduce too direct heat transfer between sample and assembly, see also ref. 5).

For the present mounting of Fig. 14b, the thermal relaxation rate may be roughly inferred from Fig. 16b to be in the order of 0.1 s. This is not too far from the ideal value $RC \lesssim 1/\omega$. With the (approximately known) value for the samples heat capacity, the observed relaxation time leads to an effective value for the thermal resistance $R \sim 7 \cdot 10^3$ K/W. With the roughly known contacting area of the sample, the contact resistance $2r_B/A$ (eq.(29)) is expected to be in the order of only 30 K/W. The large value for the thermal resistance cannot be explained from eq.(29) for reasonable values of the thickness, unless the bulk thermal conductivity is assumed to be lower by one or two orders of magnitude than the values found from ref. 39. A thickness of the bonding layer in the μm or even 10 μm range may be plausible because the surface of the samples was somewhat wavy. It may be concluded therefore that a much lower thermal resistance should in principle be attainable with a bonding contact and the high value obtained by bonding the sample around its perifery, yielding the approximately desired resistance, with satisfactory shielding against leakage, is somewhat fortunate. A possible cause might be a reduction of the effective area by void formation when the dilutant evaporates during drying and curing, which is known to be important for GE varnish[39], or a poor adhesion. It seems not very dependent on whether graphite is mixed in the varnish. Also a relaxation time of the same order of magnitude was observed when the sample was soldered.

The amplitude of the temperature rise due to a power absorption P_{abs} will be in the order of $T_b^\omega \sim \frac{1}{2} RP_{abs}$, the factor $\frac{1}{2}$ approximately taking into account the finite ωRC value. For the current biassed bolometer, with its sensitivity given by (27), the amplitude of the ac voltage becomes $v_{ac}^\omega = VRP_{abs}$ (R thermal resistance, V dc voltage across the bolometer.) The maximum dc voltage was about 400 mV before self-heating becomes important. In practice, it was ∿200 mV. (The maximum value will depend somewhat on the thickness and cooling of the bolometer). The observed value of v_{ac}^ω in Bi at zero field was typically in the order of ∿1 mV for the average FIR laser line. Bolometer signal levels are useful down to 1 μV (for proper choise of lock-in time constant ∿1 s) so that the signal level is greatly above bolometer noise and the bolometer is sensitive enough to allow detection of small changes in absorption. A signal level of 1 mV corresponds to a value $P_{abs} \sim 0.7$ μW (and a temperature rise $\sim 2\ 10^{-3}$K). The power level from the laser is in the order of $\gtrsim 1$ mW. The absorption of Bi at zero field is in the order of 1% or less[45] (below the plasmafrequency). At least an order of magnitude in power is therefore lost, presumably in the lightpipe system. Such high losses for transport through rather narrow (9 mm) lightpipes containing collecting cones (2 in the present system) and joints is not unusual in the far infrared.

The magnetoresistance of A&B resistors is about 0.02 %/T up to fields of 15 T[46] (Somewhat lower even in low fields (<3 T)). This may add a background to the apparent absorption signal. In the present work, this background is negligible compared to the absorption signals. It may become important, depending on the width of the absorption variations and their strength. In principle, it is possible to cancel the magnetoresistance by taking a second nearly identical resistor in a suitable bridge circuit. The normalizing method will also reduce the effect of magnetoresistance. Some care is required as far as the location of the reference bolometer is concerned. Several locations have been tested in the present system A location too close to the sample e.g. somewhere at the lower end of the lightpipe, the bending mirror etc. was unsatisfactory for Bi as the sample, because then the reference signal depended strongly on the (strongly varying) reflectivity of the Bi sample. A convenient mounting was as shown in Fig. 14b. A small hole drilled in the lightpipe directed a small fraction of the input beam to the detector. The reference bolometer was thus directly heated and had a shorter respons time than the signal bolometer. Indirect heating via a dummy sample to which the reference bolometer is mounted would be easily feasible and would make the respons times of signal and reference bolometer more comparable and thus improve the effectiveness of ratioing.

A field modulation technique to better resolve fine structure or to detect easier small changes in a high background is quite classical for the present sort of experiments. For this purpose, a superconducting modulation coil[47] was standard installed in the 7 T coil. A typical derivative trace is shown in Fig. 18, clearly showing the improved resolution of the fine structure. Field modulation was usually not employed however. It is somewhat more complicated than a direct absorption measurement, especially for calorimetric detection. The modulation amplitude in practice was limited by eddy current heating effects to $\lesssim 10^{-2}$ T at $\sim$ 10 Hz. As the derivative signal is proportional to the modulation amplitude, the usefulness of field modulation (and ultimate sensitivity) is strongly dependent on the sharpness of the structure in the absorption.

The intensity distribution of the FIR radiation over the sample area is strongly nonuniform when the radiation is transported by a lightpipe.

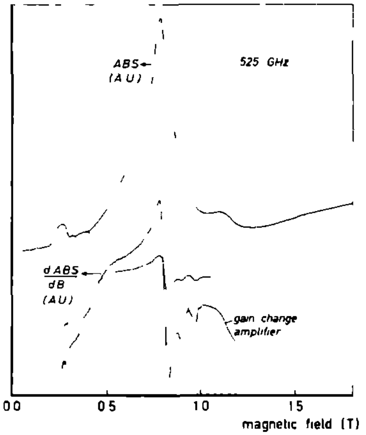


Fig. 18. Direct absorption recording of a polycrystalline Bi-sample as a function of magnetic field and its derivative trace obtained by field modulation.

This can be nicely demonstrated with a polycrystalline Bi sample (see Fig.19). The three traces (a), (b) and (c) of Fig. 19 were made shortly after one another. For each trace however, the (lightpipe) connection, guiding the radiation from the laser to the insert entrance window was independently adjusted and modified, resulting in (arbitrary) changes in the intensity distribution at the lower end of the insert light pipe. The variation of the relative intensities of the two peaks at ∿0.3 and ∿1.2 T in Fig. 19 is believed to result from predominant illumination of different crystallites. The low field peak can be identified as a dielectric anomaly occurring for

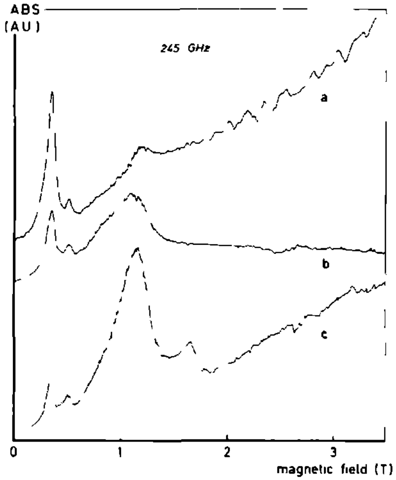


Fig. 19. Three absorption traces obtained for a polycrystalline Bi-sample as a function of magnetic field. The lightpipe system was adjusted in each case and different crystallites are apparently illuminated, showing the nonuniformity of the beam across the sample area.

trigonal plane crystallites. The high field peak is not identified, but is certainly not due to trigonal plane faces (see chapter III).

An experimental problem is sometimes posed by the proper FIR line selection when more FIR lines can be generated by the same pump transition. If only few FIR cavity modes oscillate, line selection is easy. When many modes are present, it is often more complicated and can easily lead to confusion. Fig. 20 illustrates this problem, where the absorption spectrum of a single crystal Bi sample appears to consist of a superposition of the spectra from two different lines. This can be seen from comparison with the single frequency spectra shown in chapter III. For this particular case, the laser apparently oscillates at the two frequencies 638 and 1180 GHz (471 and 251 μm) simultaneously. The FIR gas is pumped by the 10R38 line and can, besides the wavelengths mentioned, also oscillate on a 163 μm line (which is, in fact, the strongest). Simultaneous oscillation of the two lines in Fig. 20 is in fact somewhat surprising as they are assigned[25] to result from the same upper laser level. The upper level is however slightly split (nearly doubly degenerate). Moreover, assignment for the 251 μm line is not completely unambiguous[25]. It seems also that the wavelength values reported in the literature differ somewhat (in the 1% range) indicating that they actually may consist of two different lines. In this case (and other analogous cases), no attempt was made to accurately measure the wavelength, as differences in the 1% range are not significant for the purposes in the present work. With some care in finding a proper cavity setting for the FIR laser (tuning to a proper mode), it was always found possible to have the laser oscillating on one line at a time.

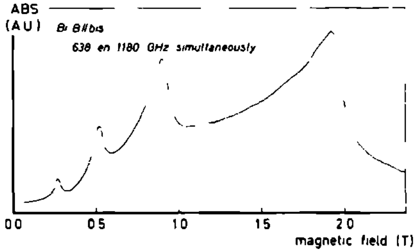


Fig. 20. Absorption signal as a function of magnetic field B for a single crystal Bi-sample (B // bissectrix axis) with the laser oscillating on two lines simultaneously.

Using filters has proved to be useful for line selection if two (or more) lines are present which have considerably differing wavelengths. A piece of carton was found to be an effective low pass filter, while single metal mesh of suitable mesh constant is useful as a high pass filter. When wavelengths are not much apart, selection is more awkward. In principle, the use of an external tunable interferometer (e.g. Fabry Perot) is very helpful for such cases[48], but has not been used for this purpose in the present work.

5. Use of a cavity.

5 i . General considerations.

As mentioned in the introduction to this chapter, use of cavities is very common in microwave spectroscopy on highly reflecting samples (metals). In the far infrared, reflection measurements are complicated if the radiation cannot be propagated freely to the sample. Free propagation is generally not possible in FIR magnetospectroscopy as the available dimensions are too small, the distances over which propagation has to occur too large and consequently diffraction effects too severe. Single mode waveguides and other components, which would enable controlled manipulation of the beam are not available. The very ingenious, but rather complex reflection spectrometer due to Allen et al.[4], using a Fabry Perot cavity and a series Fresnel aperture waveguide should be mentioned here, however.

The high reflectivity problem is circumvented by the direct absorption measurement method. If radiation shielding problems can be sufficiently solved and the bolometers are sensitive enough, then there seems no much need to use a cavity in addition to the absorption measurement method. There are however several good reasons to consider operation in a cavity. The simple reason of increased power absorption will apply in some cases of interest, e.g. when the expected changes in the absorption are very small, in addition to a rather small over-all absorption. The signal changes may then approach or exceed the bolometer noise level, or the background due to magnetoresistance. Even if signals are above noise without a cavity, an increased power absorption will allow higher resolution and observation of more details. Field modulation with modulation amplitude much smaller than the signal width also requires a signal strength considerably above the noise level.

There is another, more general reason, which might be even more important. If a cavity is used, only a resonant, well defined mode is allowed in the cavity, the power of which is largely absorbed by the sample. Stray and leakage radiation at the same time can be reduced. As discussed in the previous section, the discrimination of absorption signals against leakage signals is of crucial importance for succesful use of the calorimetric technique. Moreover, the tight bonding requirements of the sample against the housing, which is undesired, might be relaxed.

If A denotes the absorptivity of the bare sample and A_{FP} the effective absorptivity of the sample when it forms one mirror of a Fabry Perot cavity, the maximum value of A_{FP}, with the FP tuned at resonance is found from (1a) as:

$$A_{FP} = A \frac{1-R_1-A_1}{(1-\sqrt{R_1}\sqrt{1-A})^2} \quad (\equiv A \frac{T}{(1-R)^2}) \tag{34}$$

R_1 and A_1 are the reflectivity and absorptivity of the entrance mirror. As expected, for negligible A_1 and $1-A \sim R_1$, the factor multiplying A in (34) is approximately the Finesse of the cavity. Eq. (34) is illustrated in Fig. 21, where the solid lines represent the ideal case with an entrance mirror having no absorption losses ($A_1=0$). The dashed curve is a more realistic example with some finite absorption losses in the entrance mirror included. This has only the (trivial) effect of suppressing the absorption in one mirror because part of the power is absorbed by the other mirror.

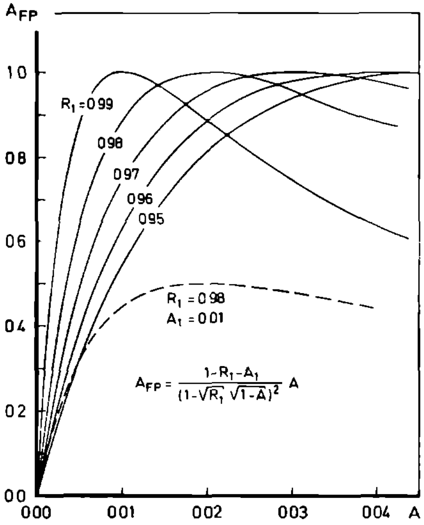


Fig. 21. Calculated effective absorptivity A_{FP} of a sample which is part of a Fabry Perot cavity as a function of its bare absorptivity A. The different curves correspond to different values of the reflectivity and absorptivity of the second (entrance) mirror.

Discussion is most instructive for the ideal case (but remains essentially unaffected for the realistic case). The absorption A_{FP} is maximal for the condition $1-R_1 = A$. Under this condition, all incoming power is coupled into the cavity and absorbed by it, no power is reflected. This is a feature familiar from microwave cavities[49]. These are usually treated in terms of Q values, where an unloaded Q factor Q_a is defined, which represents the intrinsic absorption losses of the cavity. ($Q_a \sim 1/A$). An external Q, Q_{ext} (or coupling Q) represents the losses due to reflection by the coupling hole ($Q_{ext} \sim 1/(1-R)$). Perfect match is obtained (all power absorbed in the cavity) for $Q_a = Q_{ext}$. If $A < 1-R$, the resultant Q (loaded Q) is limited by the coupling losses and the cavity is called overcoupled ($Q_a > Q_{ext}$). For the reverse inequality $A > 1-R$ ($Q_a < Q_{ext}$), the cavity is called undercoupled. Note from Fig. 21 that A_{FP} is stationary with respect to A near the matching condition (i.e. A_{FP} is independent of A to first order), which is also known from microwave theory. It should be noted that in equations describing the power flow for empty cavities, Q 's can be replaced by F 's as well[49] ($Q \sim (2d/\lambda)F$, (6)). For microwave cavities there is usually no difference between Q and F as typical cavity dimensions are in the order of a wavelength and a Finesse is not defined here. Power flow (and absorption) in empty cavities is completely controlled by F, not by Q.

For the relevant experiments, interest is in changes in A, the cavity serving to increase the over-all absorption. For this purpose, the cavity should clearly be overcoupled i.e. the effective reflectivity of the entrance mirror must satisfy $R < 1-A$. The power absorptivity of the sample is then increased by a factor $\sim F$. Operation near the power matching condition must be avoided since then A_{FP} is not sensitive to changes in A. Operation in the undercoupled region also is not very sensitive.

5 ii. Test of a cavity at low temperature.

In the experimental apparatus, the option was available to work in a hemispherical FP cavity (see Fig. 22). For this purpose, a demountable piece of lightpipe could be replaced by a spherical mirror of the lens+mesh type (see section 2). This mirror could be translated from outside the cryostat by activating the differential screw mechanism. The lack of alignment requirements greatly support the choise for a hemispherical instead of a plane parallel cavity. Alignment cannot be accomplished very conveniently when working in a cryostat.

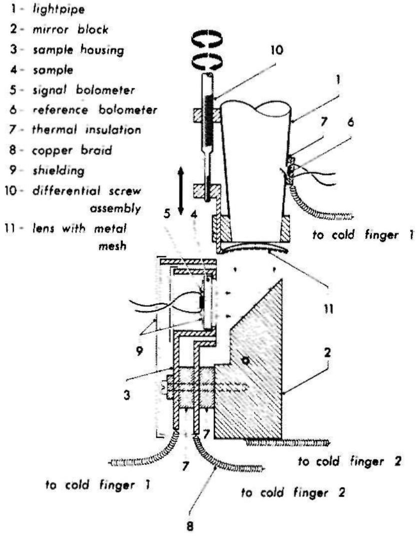
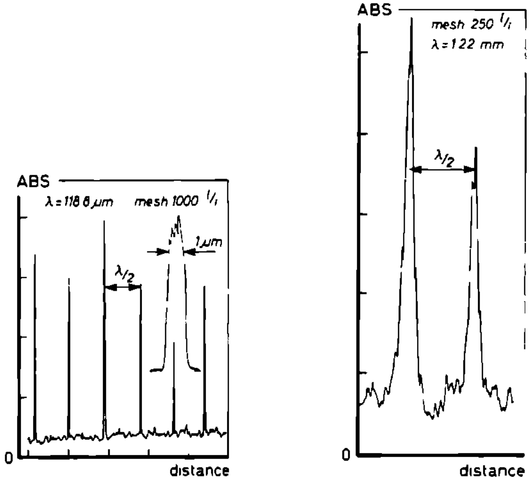


Fig. 22. Detail of the low temperature system with Fabry Perot option installed. (compare with Fig. 14(b)).

There are two possible choises for the mirror location, either very close to the sample or at considerable distance from it with a bending mirror in the cavity. As mentioned previously, the quality factor of a cavity (proportional to distance) is not of direct relevance for the present purpose, only the number of roundtrips ($\sim$F) is important (as expected from common sense wisdom). Thus there is no reason to increase Q merely by increasing the distance. Yet, a choise was made for the long distance configuration of Fig. 22. This configuration is much more easy to realize in practice (at least for the sample in vertical position). It was also expected that in this configuration the cavity could be more efficiently excited. (A bend in the lightpipe before the FP will give additional distortion of the incoming intensity distribution pattern). Also the diameter of the standing wave pattern inside the cavity becomes small (see eq. (13)) and poorly matched to the diameter of the incoming beam. At small distances, many modes are possible, which may start to overlap, although it is not clear to what extent this influences the usefulness of the cavity. Finally, diffraction losses at large distance are not larger than those at short distances for the lowest order modes.

The higher Q, associated with the longer distance is still compatible with the source frequency stability. From (1a") it is seen that intensity fluctuations inside the cavity due to source frequency fluctuations are given by $\Delta I/I \approx Q^2(\omega/\omega_0)^2$. For typical distances used (15-18 mm) and good Finesse (F $\sim$ 100), the best value for Q at the shorter wavelengths is $Q \lesssim 10^4$. The frequency stability of the laser is at least as good as 10^{-7}, so that $\Delta I/I < 10^{-6}$. At any distance, the Q value is high enough, and, more important, expected changes in Q at relevant conditions are so small that the Q dependence of the resonance frequency (not discussed up to now) can be completely neglected.

The operation of the low temperature Fabry Perot cavity, in conjunction with the calorimetric detection method, has been tested in practice. For the "sample" a dummy was used consisting of a polished copper plate. Some results are shown in Fig. 23. The results are very comparable with typical room temperature transmission FP data, described in section 2. The Finesse at 1.22 mm (Fig. 23b) is apparently somewhat low, but can be partly attributed to an imperfect mirror drive. Diffraction losses are not completely negligible also at this wavelength for a non-aligned system. The data of Fig. 23a have actually been taken taken from an earlier test set up which was nominally similar to the system of Fig. 22.



(a) *(b)*

Fig. 23. Typical results obtained with the low temperature system with Fabry Perot option, combined with calorimetric detection.

5 iii. Excitation of cavities.

A crucial and difficult problem, the excitation of the cavity, has not been discussed up to now. This is the problem of converting the input beam into the resonant mode of a cavity. With the lens excitation of Fig. 22, which was the method employed by Allen et al.[4], the focal length and the beam diameter may be optimized. Transformation of Gaussian beams by lenses is well defined[2]. The beam diameter outside the cavity should match the beam diameter inside the cavity. The wavefront curvature of a transmitted beam should match the curvature of the standing wave pattern. Assuming a nearly plane wave incident on the lens (or a beam waist width of the incoming Gaussian beam), with the known beam diameters at the two cavity mirrors (eq. (13)), the focal length can be found (using eqs. (60) and (61) of ref. 2). With the concave curvature of the lens fixed by the chosen cavity configuration, the convex curvature follows then from the lens formula. As an aid in beam diameter matching at the lens, the distance of the cavity mirrors can be properly chosen. For this reason a cavity was used in ref. 4 with $d \lesssim R$, i.e. close to the stability limit. The beam diameter at the lens can then be made arbitrarily large and the diameter at the sample small (eq. (13)).

We have chosen for a distance well below the radius of curvature, as this is easier for alignment (in combination with the larger diffraction losses for $d \rightarrow R$.) The minimum distance for the configuration of Fig. 22 is about 15 mm (determined essentially by the lightpipe diameter). The radius of the curved mirror is 19 mm. It is evident that the cavity can only be very inefficiently excited from a lightpipe as the beam front is rather arbitrary and uncontrolled. Calculation of optimum lens or cavity parameters therefore has not much sense.

To keep a well defined wave front, a Fresnel aperture waveguide was used previously to guide the radiation[4]. This will indeed be effective for improving the excitation, but such waveguide has very high losses by itself. Apart from the high losses, such guide is rather complicated and needs alignment with the laser beam. The usefulness of the presently used lightpipe and cavity must be found from practice. A cone was used at the lower end of the lightpipe to decrease the beam diameter and match more properly the cavity beam diameter. At the same time however, the cone will go at some additional cost of beam quality. It should be mentioned that for lightpipe properties and collecting cone properties, there should not be relied on calculations based on geometrical optics. Diffraction effects are clearly important over the distances involved (see e.g. eq. (11)), although strongly dependent on wavelength, and may cause some "self-shaping" of the beam.

For phenomenological calculations, one might define a fraction α of the incoming power P_0, which is converted into some resonant cavity mode. The power absorbed due to this resonant fraction $P_r = \alpha P_0 \cdot A_{FP}$. With the rough assumption that the rest of the power is nonresonantly transmitted and strikes the second mirror once before walking out of the cavity, the power absorbed due to this nonresonant fraction is $P_{nr} = (1-\alpha)\ P_0 \cdot T \cdot A$. T is the transmissivity of the entrance mirror, $T \equiv 1-R_1-A_1$. If multiple reflections of the "wrong" intensity distribution were taken into account, the nonresonant power absorption would be lower by at most a factor of ~ 4 by destructive interference. This may be inferred from inspection of the Airy formula (1a) for transmitted intensity for a cavity out of resonance, or equivalently, for radiation impinging in a "wrong" direction. Possible off-normal, but resonant directions (the "rings" of a FP) should be thought to be contained already in the resonant fraction α. Without any reflection at all, the nonresonant power absorption by the sample is zero. The above estimate for P_{nr} is thus an order of magnitude estimate. If (interfering) multiple reflections (for nonresonant radiation) are important, the transmitted power is reflected back from the cavity in the lightpipe. Otherwise the radiation walks out of the cavity and may contribute to stray and leakage radiation, apart from sample heating. When the bare sample was subjected to the same radiation field, without FP, the absorbed power would be $\sim P_b = A \cdot I$.

The ratio of resonant to nonresonant absorption is given by, using (34):

$$\frac{P_r}{P_{nr}} = \frac{\alpha}{1-\alpha} \cdot \frac{1}{(1-R)^2} \tag{35}$$

(The effect of the resonant fraction is enhanced by a factor $1/(1-R) \sim F$, while the nonresonant fraction is suppressed by a factor 1-R; this result is trivial for an FP having mirrors of equal reflectivity R). Observation of an FP pattern will be possible if only a fraction $\alpha \simeq (1-R)^2 \simeq (\pi/F)^2$ (making $P_r/P_{nr} \sim 1$) is resonantly coupled in. For a typical experimental value for F, this corresponds to $\alpha \sim 3 \cdot 10^{-3}$! Experimentally, a reasonable background was always present in the low temperature absorption measurements (typically $\gtrsim 10\%$, see Fig. 23), which is much higher than expected for the signal between resonance peaks for resonant radiation (signal at resonance = $4/(1-R)^2$ × signal out resonance, see eq. (1a)). Assuming (35), a typical measured value for F and typical peak to background ratio, a value $\alpha \sim 3\%$ is obtained. In transmission measurements at room temperature (see section 2), using the same FP geometry and with the radiation transported over a considerable length of lightpipe, similar to the one used in the dewar, the background was usually negligible and consistent with the limiting background for an ideally excited FP. (It is simple to show that (35) holds also for transmission). Similar room temperature tests were performed with one 90° bend in the lightpipe (via a 90° bending mirror) and an increase in background level occurred, but was still considerably lower than 10%.

The experimental system of Fig. 22 itself was also tested in a transmission configuration, both at room temperature and at low temperature. For this purpose, the dummy "sample" was replaced by a mesh mirror, for convenience glued on a flat TPX substrate. At both temperatures, the Fabry Perot patterns were similar and showed a considerable background now, comparable (and sometimes even larger) to the (absorption) data displayed in Fig. 23. A possible explanation for this increased background could be the larger distortion of the incoming beam (effectively decreasing α). There exists also indication that radiation leaking out between the lightpipe end and the lens contributes to the signal.

It is important to compare the absorption signal of resonant absorption with the absorption which would have been obtained without FP. Their ratio is given by:

$$P_r/P_b = \alpha \cdot A_{FP}/A = \alpha \cdot \frac{T}{(1-R)^2} \qquad (36)$$

within the assumptions made above. Using typical values estimated for the cavities tested (which are not optimal), a value for α of a few % (very roughly π/F) is required to have at least as much absorption with an FP than without. Practical values for α were above estimated to be a few percent. This indicates that with the present system, the FP is not effective in increasing the absorption strength, the practical status being that roughly the same absorption is obtained with or without a FP. As outlined in the introduction to this chapter, even in this case a FP might be highly useful.

The cone plus copper mirror system, employing a coupling hole, of which a prototype version was tested in a special arrangement in the transmission configuration (see section 2) could not be incorporated in the present low temperature apparatus. It is expected however, that this system can be superior to the lens sytem in several respects. Without calculations, it may be taken as a matter of fact from microwave experiments, using properly scaled versions of the same geometry[6], that excitation through the coupling hole from a single mode waveguide is highly efficient, close to 100 %. For the far infrared, the problem is then just shifted to excitation of a single mode in the short piece of waveguide just before the coupling hole (see Fig. 3), which suffers from the same sort of problems as exciting the Fabry Perot by a lens. Also absorption losses in the lower end of the cone and waveguide are much more severe in the far infrared than in the microwave. At the shorter wavelengths of the FIR, the waveguide will necessarily be oversized still for reasons of manufacturing. The absorption losses however will favour effective propagation of lower order modes.

The effective performance can be reliably obtained from practice only. In the test experiments, described in section 2, it was found that the copper mirror system was at least as effective as the lens system (but may compare favourably). This was true at long wavelengths, where the waveguide was designed to be nearly single mode. It was also true at short wavelengths, where the waveguide was overmoded. Short wavelength operation seems to be limited by the reflection losses through the coupling hole, which cannot easily be made small enough (and the diaphragm thin enough). Apart from the possibly favourable comparison with the lens system, the copper mirror has one more important virtue over the lens system. The fraction of power coupled in the waveguide should be assumed to couple very efficiently[6] into the cavity. The rest of the power however is reflected back into the lightpipe from the cone or absorbed in it, but does not enter the cavity and so does not contribute to leakage radiation to the absorption bolometer. In addition, from Fig. 21, it follows that the absorptive losses in the input mirror should be as low as possible. For a high quality, polished copper mirror, which may be gold-plated, the absorption is expected to be much less than for a mesh.

5 iv. Summarizing remarks.

A general optimization of a FP cavity cannot be given because it depends on the application. As optimization depends strongly on wavelength, the more optimized the system is, the less large is the wavelength range where it can be used. Some general considerations can be summarized. Clearly, the F value should be as high as possible, but the cavity must always be somewhat overcoupled. Overcoupling of the cavity will in practice always be fulfilled as the samples of interest will have high reflection in the far infrared, so that the other mirror will have a lower effective reflectivity because it has coupling losses in addition to its absorption losses. The absorption losses of the input mirror should be as low as possible. Practically, the absorption losses cannot be made much lower than the reflection losses for high F cavities (as the F will eventually be limited by absorption losses).

In the experiments on Bi, described in chapter III, no cavity has been used. Most of the relevant absorption features were very strong here (A >> 1%, perhaps approaching 100% in some cases). Not only is a cavity useless here, but moreover it would even distort the signals as the measured absorption depends highly nonlinear on the sample absorption (see Fig. 21).

Recently, a new low temperature apparatus[50] has become available. It contains several features not possible with the present system. The sample can be aligned with the magnetic field from top of the dewar. Also the sample can be rotated to align a crystal direction with the magnetic field. The geometry of the FP is essentially unchanged, but a novel and original mirror drive mechanism[50] has been employed. Both a lens/mesh mirror or a cone/copper mirror with coupling hole can be installed. Preliminary tests of this system, especially regarding the Fabry Perot performance, seem promising.

REFERENCES

1. G.D. Boyd and J.P. Gordon, Bell Sys. Tech. J. 40, 489 (1961).
2. H. Kogelnik and T. Li, Proc. IEEE 54, 1312 (1966).
3. R. Ulrich, K.F. Renk and L. Genzel, IEEE Trans. Microwave Theory Tech. MTT-11, 363 (1963).
4. S.J. Allen, Jr., L.W. Rupp, Jr. and P.H. Schmidt, Phys. Rev. B 7, 5121 (1973).
5. P. Goy and B. Castaing, Phys. Rev. B 7, 4409 (1973).
6. R.W. Zimmerer, IEEE Trans. Microwave Theory Tech. MTT-11, 371 (1963).
7. T.Y. Chang and T.J. Bridges, Opt. Commun. 1, 423 (1970).
8. T.Y. Chang, IEEE Trans. Microwave Theory Tech. MTT-22, 983 (1974).
9. D.T. Hodges, J.R. Tucker and T.S. Hartwick, Infrared Phys. 16, 175 (1976).
10. D.T. Hodges, Infrared Phys. 18, 375 (1978).
11. J.-M. Lourtioz, R. Adde, D. Bouchon and J. Pontnau, Rev. Phys. Appl. 14, 323 (1979).
12. For a review, see: A.J. DeMaria, Proc. IEEE 61, 731 (1973) and references herein.
13. D.T. Hodges, F.B. Foote and R.D. Reel, IEEE J. Quantum Electron. QE-13, 491 (1977).
14. J.O. Henningsen and H.G. Jensen, IEEE J. Quantum Electron. QE-11, 248 (1975).
15. T.A. De Temple and E.J. Danielewicz, IEEE J. Quantum Electron. QE-12, 40 (1976).
16. H.J.A. Bluyssen and A.F. van Etteger, unpublished.
17. W.W. Rigrod, J. Appl. Phys. 36, 2487 (1965).
18. D.T. Hodges and T.S. Hartwick, Appl. Phys. Lett. 17, 249 (1973).
19. E.A.J. Marcatili and R.A. Schmelzer, Bell Sys. Tech. J. 40, 489 (1961).
20. J. Degnan, Appl. Phys. 11, 1 (1976).
21. E.J. Danielewicz, T.K. Plant and T.A. De Temple, Opt. Commun. 13, 366 (1975).
22. G. Busse, E. Basel and A. Pfaller, Appl. Phys. 12, 387 (1977).
23. G. Busse, E. Basel and A. Pfaller, Appl. Phys. Lett. 31, 194 (1977).
24. G. Busse, Infrared Phys. 19, 175 (1979).
25. J.O. Henningsen, IEEE J. Quantum Electron. QE-13, 435 (1977).
26. J. Degnan, J. Appl. Phys. 45, 257 (1974).
27. J.-M. Lourtioz, J. Pontnau and F. Julien, Infrared Phys. 20, 231 (1980)

28. A.F. van Etteger, unpublished.

29. H.J.A. Bluyssen, R.E. McIntosh, A.F. van Etteger and P. Wyder, IEEE J. Quantum Electron. QE-11, 341 (1975); H.J.A. Bluyssen, R.E. McIntosh, A.F. van Etteger and P. Wyder, Infrared Phys. 16, 183 (1976).

30. R.J. Temkin, IEEE J. Quantum Electron. QE-13, 450 (1977).

31. D. Seligson, M. Ducloy, J.R. Leite, A. Sanchez and M.S. Feld, IEEE J. Quantum Electron. QE-13, 468 (1977).

32. G.A. Koepf and K. Smith, IEEE J. Quantum Electron. QE-14, 333 (1978).

33. J. Heppner and U. Hübner, IEEE J. Quantum Electron. QE-16, 1093 (1980).

34. Z. Drozdowicz, R.J. Temkin and B. Lax, IEEE J. Quantum Electron. QE-15, 170 (1979).

35. Z. Drozdowicz, R.J. Temkin and B. Lax, IEEE J. Quantum Electron. QE-15, 865 (1979).

36. M. Inguscio, A Moretti and F. Strumia, Opt. Commun. 30, 355 (1979).

37. J. Heppner and C.O. Weiss, Opt. Commun. 21, 324 (1977).

38. G. Duxbury and H. Herman, J. Phys. B 11, 935 (1978).

39. A.C. Anderson and R.E. Petersen, Cryogenics 10, 430 (1970).

40. H.S. Carslaw and J.C. Jaeger, *Conduction of heat in solids,* 2nd edition (Clarendon Press, Oxford, 1962).

41. D.L. Carter and J.C. Picard, Solid State Commun. 5, 719 (1967).

42. M. von Ortenberg in *Infrared and Millimeter Waves,* Vol. 3, K.J. Button, ed. (Academic, New York, 1980), Ch. 6.

43. R.L. Blewitt and A.J. Sievers, J. Low Temp. Phys. 13, 617 (1973).

44. G.K. White, *Experimental techniques in low temperature physics,* 2nd edition (Clarendon Press, Oxford, 1968).

45. C. Nanney, Phys. Rev. 129, 109 (1963).

46. L.J. Neuringer and Y. Shapira, Rev. Sci. Instr. 40, 1314 (1969).

47. J.C. Maan, unpublished.

48. J.C. Maan, Thesis, Nijmegen, 1979.

49. J.C. Slater, *Microwave Electronics* (D. van Nostrand Company, Princeton, 1959), Ch. 5.

50. J.G.H. Hermsen, unpublished.

CHAPTER III. FAR-INFRARED BEHAVIOUR OF THE MAGNETOPLASMA EFFECTS IN Bi IN QUANTIZING FIELDS.

ABSTRACT

The far-infrared absorption of Bi is investigated as a function of magnetic field at a fixed frequency. The experiments were carried out under conditions where $\omega \sim \omega_c \sim \omega_p \sim E_F/\hbar$, where ω_c the cyclotron frequency of the relevant carriers, ω_p the plasma frequency and E_F the Fermi energy. This condition is satisfied in Bi in fields of a few Tesla in the far infrared. A detailed study is possible by utilizing the full frequency range which can be covered by an optically pumped far-infrared laser, going from ∿ 250 GHz, where typical low frequency, classical behaviour occurs, up to ∿ 7000 GHz, which is well above the plasma frequency. In this frequency range, the magnetoplasma absorption features occur in a field range near the quantum limit for the relevant set of carriers, which means that only one Landau level is occupied below the Fermi level. Measurements were always taken on a trigonal plane sample in the Voigt configuration, with the magnetic field either along a binary or bisectrix axis.

The fundamental quantum cyclotron resonance line shape reflects details of the nonparabolic bandstructure of Bi. Structure due to $k_B \neq 0$ transitions (k_B momentum along magnetic field) can be clearly identified. Positions of hybrid resonances and lineshapes of dielectric anomalies are strongly influenced by quantum effects. This can be attributed primarily to a carrier redistribution over different pockets. Very complicated absorption spectra occur for frequencies near the plasma frequency, where a transition between low-and high-frequency limits takes place, which is at the same time accompanied by quantum effects.

For analyzing the data, the absorption is calculated in three different ways. First, a purely classical model is used, which is extended phenomenologically in the second model by including the field dependent carrier redistribution. In the last model, a more rigorous quantummechanical calculation is carried out, which takes the bandstructure properly into account. A comparison of the latter calculations with the data has been used as a test

for the velocity matrix elements of the two band model, describing the nonparabolic bandstructure of Bi. As an approximation, a local relation between currents and fields was always assumed.

For comparison with the effects observed in Bi, some data were also taken on Sb, which is very similar to Bi, except for a two orders of magnitude higher carrier concentration and effective masses. High-frequency - or quantum effects are negligible here, but the electrodynamical conditions are expected to be much more nonlocal as compared to Bi. Nevertheless, the rich structure in the high-field absorption in Sb, observed experimentally, could still be unambiguously related to calculated spectra from a strictly local approximation. These phenomena have not been investigated previously in Sb or interpreted as such.

1. Introduction.

The semimetal Bi is by now a thoroughly investigated material. Being a semimetal, it has all the characteristic properties of a real metal, yet owing to the low carrier density, as well as to the low effective masses, it can be studied under typical conditions, which are experimentally inaccessible for real metals. It therefore has generally served in the past as a model system , in which a great variety of phenomena have been studied. In modest magnetic fields, some of the carrier pockets reach the extreme quantum limit, which means that only one Landau level is occupied below the Fermi level.

The Fermi surface has been studied extensively by the usual methods and is well established[1]. The electron Fermi surface consists of three elongated, nearly ellipsoidal pockets, located at the L-points in the Brillouin zone and one hole ellipsoid of revolution at the T-point in the zone. Because of the small band gap between valence and conduction bands at the L-point, the energy dispersion relation for the electrons is strongly nonparabolic[2].

Both for investigating the energy spectrum of the carriers, but primarily to obtain an understanding of the behaviour of a dense solid state plasma, extensive spectroscopic investigations do exist over a broad range of the spectrum, usually in combination with an applied dc magnetic field. The behaviour of the plasma depends strongly on whether the spatial dispersion of the high-frequency fields is important and therefore the observed phenomena can be generally classified in local and non-local effects. At low frequencies and low magnetic fields, $\omega_c \lesssim\!< \omega$ where ω_c the cyclotron frequency and ω the applied frequency, nonlocal effects prevail (Azbel-Kaner cyclotron resonance[3,4]). In the same frequency range but at higher fields ($\omega_c >> \omega$), the high-frequency fields can penetrate deeper into the metal because of the low carrier density and the conditions become local. The collective behaviour of the electron gas manifests itself then in the propagating modes of electromagnetic radiation which are sustained by the plasma. (Alfvèn waves[5]).

In the intermediate field range ($\omega_c \sim \omega$), the conditions are predominantly local, but the behaviour of an anisotropic multicomponent plasma such as Bi is very complex[6]. This case was investigated and analyzed for the first time in the authorative work of Smith, Hebel and Buchsbaum for the Voigt configuration, k ⊥ B[7]. Local wave propagation can take place in certain regions of magnetic field which start at a resonance in the effective dielectric constant and terminate at a so called dielectric anomaly. Resonances result in general from the combined effect of two or more groups of carriers (hybrid resonance) or from a cyclotron resonance of one group of carriers. Cyclotron resonance is possible only if there are nonequivalent groups of carriers present, or if the orbit is tilted with respect to the magnetic field, but is otherwise shielded due to the Hall-effect. These kind of phenomena are very general and are also important in the magnetospectroscopy of high density, anisotropic semiconductors[8,9]. Also a nonlocal type of wave propagation, known as cyclotron waves, is observed and investigated in Bi[10-12].

It is the purpose of the present work to investigate the behaviour of the above mentioned magnetoplasma effects over a large frequency range in the far-infrared (FIR). While at the lower end of this range the experiments link up with the microwave results, two important modifications show up at higher frequencies. First, the applied frequency, as well as the cyclotron frequency of one group of carriers, become comparable to and finally exceed, the plasma frequency. This is in principle a classical effect. Secondly, for higher frequencies, the positions of the resonances and dielectric anomalies are shifted into the extreme quantum limit, so that quantum effects will be important. It is of interest to examine how these typical phenomena, which are understood in terms of classical magnetohydrodynamic plasma theories, evolve when they are brought into highly quantizing fields. For the cyclotron resonances, an intuitive understanding can be expected in terms of single transitions between Landau levels, but it is less clear for the essentially collective properties such as hybrid resonances and dielectric anomalies. It should be stressed however, that within the framework of plasma theory, the cyclotron resonances and associated dielectric anomalies also follow as a collective property of the charged gas and are in this respect similar to hybrid resonances. A thorough understanding of highly quantized systems is also of importance in connection with interesting many body phenomena which can occur here, such as electron-electron interaction[13,14] or electron-electron-phonon interaction in systems of the Bi type[15] or polar semiconductors[16].

Several magnetospectroscopic studies of Bi exist in the FIR. In the classic work of Boyle and Brailsford[17], the short wavelength region of the far infrared (FIR) ($\lambda \lesssim 60$ µm) was used, to observe the (magnetic field dependent) plasma resonance and the onset of interband transitions ($\lambda \sim 20$ µm). Azbel - Kaner cyclotron resonance in the FIR was investigated by Strom et al.[18,19] and revealed strongly the nonparabolic band structure because of the large photon energies. In the extreme quantum limit, Bi can become transparent to FIR radiation[20], which is attributed to a strongly reduced scattering rate[20]. This effect has been exploited to study the hole cyclotron resonance and effects associated with it, at fields above the extreme quantum limit of the light electrons[14,21,22]. The line shape of one dielectric anomaly, occurring well below the extreme quantum limit, has been investigated to study the frequency dependence of the electronic lifetime [23,24]. Far-infrared electron spin resonance has been observed in Fourier transform spectroscopy[25] as well as with monochromatic sources, using carcinotrons[26].

A considerable amount of literature is available on spectroscopic measurements of Bi in a highly quantized state. Of these, the infrared magnetoreflection studies[27,28] have been of particular importance for investigating the nonparabolic band. As the interband transitions, which are measured here, are sensitive to details of the bandstructure, deviations from the two band model could be established using high magnetic fields[29,30]. States at $k_B=0$ only however, dominate the experimental spectra. Cyclotron resonance has still been observed at optical frequencies ($\lambda = 0.6$ µm) in ultra-high (1000 T) pulsed magnetic fields[31].

At low frequencies, the quantum state has been investigated by measuring the dispersion relation of Alfvèn waves in high field[32,33]. These experiments can be interpreted by a classical analysis, but allowing the carrier density in the different pockets to vary, by calculating the appropriate Landau level degeneracy. Thus Fermilevel oscillations, or mass density oscillations, can be measured.

To analyze the data, the experimental results are compared with three types of theoretical calculations, which will be referred to as model I, II en III respectively for convenience . The simplest model, I, is identical to the calculations in the original work of Smith et al.[7] and is included to show clearly the effects which arise on purely classical grounds (plasma-frequency). Model II is essentially the same as model I, however the (magnetic field dependent) carrier density in the different pockets is calculated independently, taking into account the proper degeneracy of the

Landau levels. This indicates the quantum effects, which can be expected on phenomenological grounds. In the last model, III, a quantummechanical expression for the optical conductivity tensor is used, together with a theoretical expression for the velocity matrix element. In all cases, a strictly local condition has been assumed. This assumption is not rigorously justified, but deviations in terms of nonlocality will be discussed qualitatively only.

An example where local conditions are believed to break down seriously is Sb, because it is much more metallic than Bi owing to its ∿ 100 x higher carrier concentration. On the other hand, quantum effects nor classical high-frequency effects are important here at the frequencies used. The experimental absorption data show strong structure at fields above the cyclotron resonance fields of the light carriers. The positions of this structure can be unambiguously identified with the hybrid resonances and dielectric anomalies following from a classical local calculation. They are more numerous as compared to Bi, because of the more complicated Fermi surface of Sb.

Both for Bi as for Sb, a trigonal plane sample was used with B either along a binary or bisectrix axis. Unpolarized radiation was incident along the trigonal axis (Voigt configuration). In the following, the experimental procedure will be outlined first. In the next section, the theoretical framework within which the results have been analyzed, will be sketched. A detailed description of the experimental results will finally be given and discussed.

2. Experimental procedure.

Radiation was obtained from a conventional type optically pumped FIR laser. A waveguide type FIR laser with either a 25 mm diameter metal tube for long wavelengths, or a 13 mm diameter quartz tube at short wavelengths was pumped by a grating - and cavity tuned 50 W CO_2 laser. As laser gases CH_3OH, CH_3OD and incidentally HCOOH were used. The sample was mounted, facing the end of a 10 mm diameter stainless steel lightpipe which was irradiated by the laser as shown in Fig. 1. To decrease the beam diameter, the pipe ended in a small angle cone with 6 mm final diameter, followed by a 90^o bending mirror. Signals were detected calorimetrically. A thinned, 220 Ω Allen & Bradley resistor was attached at the back of the sample using

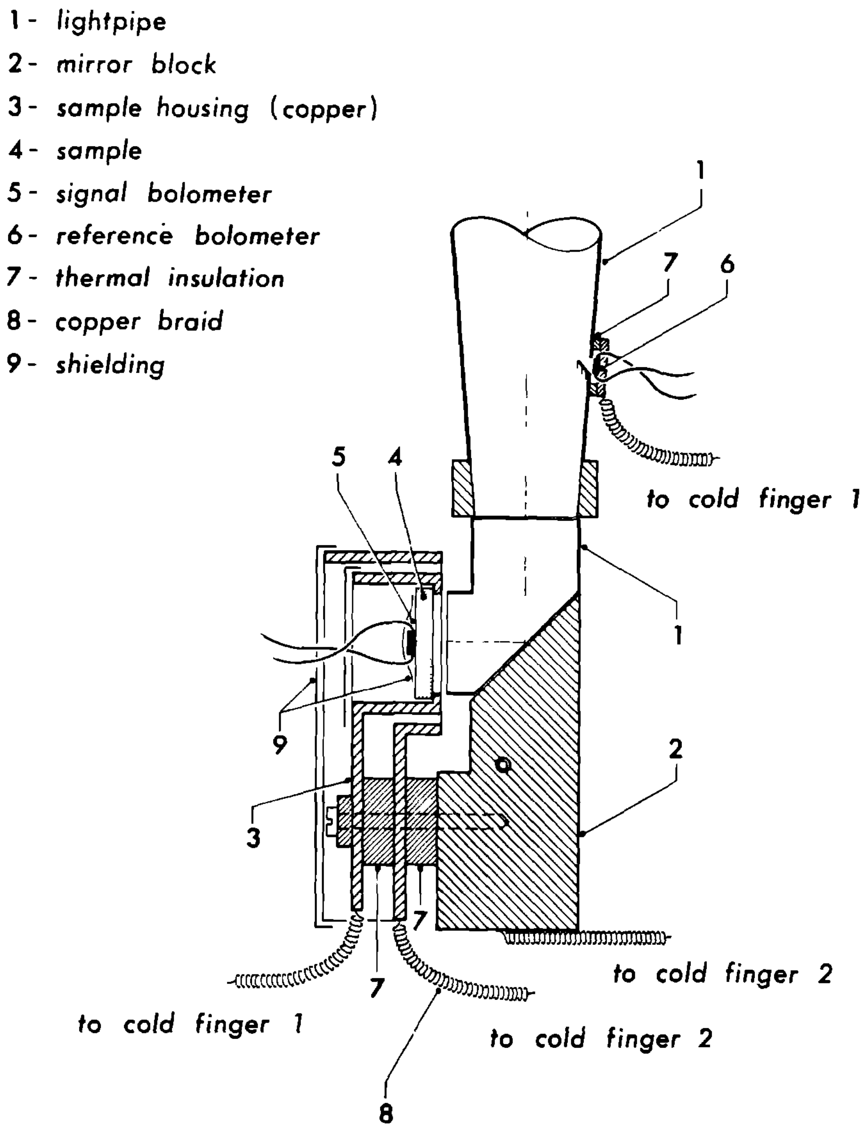


Fig. 1. Lower part of the lightpipe and sample mounting.

General Electric 7031 (GE) varnish, electrically insulated from the sample by a foil of cigarette paper. This resistor was current biased and used as a bolometer to measure the sample temperature variations due to absorption of radiation. The laser power was modulated by chopping the CO_2 laser at 11 Hz for phase sensitive detection. A second, nominally identical bolometer, was used as a reference detector and was heated by radiation leaking through a small hole drilled in the light pipe. Using the reference detector, the signal was normalized by ratioing to eliminate some laser instability ($\sim \pm 2\%$) or drift. This method also eliminates magnetoresistance of the bolometers, although this was very small ($\lesssim 5\%$).

The sample housing plus bolometer was thermally decoupled from the rest of the system by situating the assembly in a vacuum environment. The sample housing was connected to the support frame using delrin screws and hard foam spacers (Rohacel), to assure at the same time the required mechanical stiffness as well as the high degree of thermal insulation.

Two helium filled copper tubes, connected with the bath, extended into the vacuum can and served as cold fingers to cool the different parts of the assembly by connecting them to the tubes via copper braid. One tube was used to cool sample and reference detector exclusively, to avoid possible thermal cross talk between spuriously heated parts of the support frame (notably the lightpipe) and the detecting system.

The sample holder itself was machined out of one single block of copper and kept at the bath temperature. The sample was glued with its front face against a rim of the copper holder (see Fig. 1). This rim provided the reference plane with respect to the main support frame, which was aligned with the magnetic field. The magnetic field direction with respect to the sample plane is thus determined by mechanical tolerances of machining and mounting and is estimated to be in the sample front plane to within $\pm 3^{\circ}$.

The sample was carefully shielded to minimize influence of leakage radiation, which was especially severe at the longer wavelengths. The main leakage channel was formed by the spacing between the sample front plane and the rim of the mount against which the sample was pressed. This leakage could only be reduced to a negligible level by glueing the sample directly against the rim using GE varnish mixed with fine grounded graphite powder around the entire edge of the sample. The varnish layer provided a thermal resistance between the sample and its cold holder to make the thermal time constant of the sample plus bolometer about 100 ms. The thermal relaxation time of the detector alone, cooled by the sample, was about 10 ms. Signal magnitudes were generally large enough to display the bolometer signal directly on an oscilloscope, so that the relaxation times could be measured. The observed thermal relaxation times thus provided an indication of the amount of leakage radiation . As the sample was rigidly attached to the holder, some strain may be induced in it after cooling. This might indeed account in part for the rather large relaxation rate observed. Line shapes and peak positions however, were not significantly influenced by thermal recycling or mounting and remounting the samples. This indicates that possible strain effects are not important for the observed absorption features.

The samples were spark cut from nominal 5N+/6N purity single crystal material and oriented by X-ray analysis. A mark on the back of the sample indicated the main crystallographic axis in the sample plane. This mark was used to orient the sample in the desired position with reference to the main support frame. In this way the crystallographic axes could be aligned along the magnetic field with an accuracy of $\pm 3^{\circ}$.

Before mounting, the Bi sample was first etched in concentrated nitric acid and water and finally in 50-50 solution of concentrated nitric acid and glacial acetic acid[20]. The Sb sample was etched in a mixture consisting of six parts glacial acetic acid, five parts fuming nitric acid and two parts hydrofluoric acid[34]. At least 100 μm was removed and eventually a somewhat wavy, but shiny surface was obtained for both Bi and Sb. Some measurements were repeated with the sample further etched and polished, but the experimental spectra did not change significantly. Thickness of the samples was 1.5 - 2 mm.

Special care must be taken regarding the position of the reference bolometer. Clearly, the ratioing technique works most efficiently when both detectors are as identical as possible and subjected to the same conditions. However some of the signals to be described involve large changes in the reflectivity in the order of 50%, which can have a marked influence on the reference detector when it is too close to the sample , e.g. at the lower end of the lightpipe, the 90° bending mirror, or the sample holder. For these reasons, it was mounted across a small hole ($\lesssim$ 1 mm) drilled in the lightpipe some distance away from the sample. A piece of metal foil inserted in the hole, served to screen the detector against possibly reflected radiation , and coupled a little of the incident radiation onto it at the shortest wavelengths, where diffraction is inefficient. As the radiation fell directly on the bolometer, its time constant was much shorter than that of the signal detector. Mounting the reference detector on a dummy sample in a similar way as the signal bolometer, could have corrected its time constant, but in practice, ratioing worked out sufficiently well for long enough amplifier time constants.

The laser power, which varied from line to line, but was typically in the mW range, was in all cases adjusted such as to give signals well above the bolometer noise level, but was kept sufficiently low to assure a linear response. The laser radiation is highly polarized leaving the laser, but it was verified at some wavelengths that little polarization was left after being scrumbled in traversing the approximately 1.5 m total length lightpipe, which included two collecting cones. No polarizer was used in front of the sample and was omitted to maximize the useful spectral range .

The magnetic field was generated by a 7 T superconducting solenoid (although a smaller 4 T coil was mostly used in the low-field measurements). In all experiments described, the LHe-bath was pumped down to a temperature of ∿ 1.8 K.

The sensitivity of the present system was determined by the stability of the laser in combination with the effectiveness of ratioing. Under optimum conditions, changes in the absorption of ∿ 0.1% could be detected. As the absorption is generally rather low (< 1% at zero field for $\lambda >$ 100 μm), and the reflection high, a calorimetric method is clearly more sensitive than a reflection experiment, unless a multiple reflection configuration is chosen. For this reason, a strip line device was used in other work on Bi[18,19] , and degenerate semiconductors[35], while a far-infrared Fabry Perot resonator was used either in reflection[36] or in combination with calorimetric detection[37] for cyclotron resonance measurements in metals . A multiple reflection unit was not necessary in the present experiments. Moreover it clearly should be avoided in cases where reflection changes are large[24]. A calorimetric method similar to the one described here was used successfully in other work on Bi[7,15,38] and metals[37,39].

3. Theory.

An exact treatment of the electrodynamical properties of a metal surface is exceedingly difficult. An additional complication to treat the semimetals Bi and Sb is their high anisotropy and the distribution of charge carriers over several pockets of Fermi surface . Application of a strong magnetic field results in important constraints on the carrier motion and thus influences the electrodynamic properties strongly. For analyzing our results, we will first use the classical framework set out by Smith, Hebel and Buchsbaum[7]. In their work, the electrodynamic properties of the metal were derived from the high-frequency conductivity tensor, calculated in the strictly local limit, taking the anisotropy fully into account. Such treatment suffers from two important restrictions, when applied to Bi. First, nonlocal effects are clearly important, in the microwave region as well as in the far infrared and manifest themselves in the occurrence of the Azbel-Kaner effect[40] and cyclotron wave propagation effects[10-12]. Second, the Fermi-energy is very small and comparable to the Landau level spacing in not too small fields. This means that quantum effects can be important. Quantum effects are not important at microwave frequencies in low fields, but strongly influence the measured spectra in the range of fields and frequencies of the present work. These effects are especially pronounced due to the highly nonparabolic bandstructure of Bi. For this reason, we also have used a quantummechanical expression for the optical conductivity tensor.

3 i. Classical framework.

To start with, we briefly discuss the origin of the main observed features in a qualitative way following Smith et al[7].
If a local relation between currents and fields is assumed, Maxwells equations can be reduced to:

$$\vec{\nabla} \times \vec{\nabla} \times \vec{E} - \frac{\omega^2}{c^2} \vec{\vec{\varepsilon}} . \vec{E} = 0 \qquad (1)$$

where ω is the applied frequency, c speed of light and $\vec{\vec{\varepsilon}}$ the dielectric tensor given by

$$\vec{\vec{\varepsilon}} = \vec{\vec{\varepsilon}}_1 + (1/j\omega\varepsilon_o)\vec{\vec{\sigma}} \qquad (2)$$

with $\vec{\vec{\varepsilon}}_1$ the lattice dielectric tensor, ε_0 permittivity of free space and $\vec{\vec{\sigma}}$ the conductivity tensor. The replacement term $\vec{\varepsilon}_1$ should be retained, as it becomes important at high frequencies. In a classical analysis, the conductivity most easily follows from the Lorentz force equation for the velocity. Knowing the conductivity tensor, the reflection can now be calculated by inserting a plane wave $E\exp(i(\omega t-\vec{k}.\vec{r}))$ in eq. (1) and solving the resulting secular equations. In the Voigt configuration, for the magnetic field B parallel to the binary direction in Bi or Sb, two normal modes result, one polarized parallel to B (ordinary mode) and the other polarized in a plane perpendicular to B. Writing the solution in both cases as

$$k^2 = (\omega^2/c^2)\varepsilon_{eff} \qquad (3),$$

the reflection R is given by the Fresnel formula:

$$1-R = 4\ \mathrm{Re}\sqrt{\varepsilon_{eff}}\ /\ |1+\sqrt{\varepsilon_{eff}}|^2 \qquad (4)$$

When B is along the bisectrix direction, none of the tensor components is zero and the secular equation does not factorize. The rather awkward formulas which have been used to treat this case are given in ref. 7.

As the magnetoconductivity tensor for Bi is very complex because carriers of different sign are distributed over 4 (or 9 in Sb) highly anisotropic pieces of Fermi surface, it is instructive to keep in mind the solutions of (1) and (2), obtained in more simple systems. The experimentally measured quantity is 1-R, where R is the reflectivity, because the sample is sufficiently thick that transmission can be neglected in realistic cases with finite relaxation time τ. This is an important experimental condition, which means that experimentally no difference can be made between resonant absorption and relaxation damping of transmitted waves. In discussing qualitative features, we will assume $\tau = \infty$ and semi-infinite samples.

In the simplest case with one group of isotropic carriers, the dielectric constant for the ordinary polarization is independent of the magnetic field for all frequencies and has a resonance at the plasmafrequency $\omega = \omega_p$ only^2. There is no singularity at the carrier cyclotron resonance, because the orbital velocity is perpendicular to the electric field. Also in the extra ordinary polarization, with $E \perp B$, no singularity occurs at the cyclotron frequency. Because of the Hall effect, the self consistent field in the metal is elliptically polarized in the plane perpendicular to

B, but the sense of rotation is opposite to that of the carriers and no resonance occurs. The optical properties of the plasma for low frequencies are not affected by the magnetic field, unless the cyclotron frequency approaches the plasmafrequency and the plasma becomes transparent at the dielectric anomaly given by[2]

$$(\omega_c/\omega)^2 = (1-(\omega_p/\omega)^2) \tag{5}$$

This dielectric anomaly is also referred to as the (magnetic field dependent) plasma edge.

For $\omega > \omega_p$, a dielectric anomaly (defined as a zero in the dielectric constant) occurs again at a field value given by eq. (5), followed by a resonance, which is shifted from the cyclotron resonance and given by[2]

$$\omega_c/\omega = \sqrt{1-(\omega_p/\omega)^2} \tag{6}$$

The situation becomes different if the plasma consists of more than one type of carrier. As in the case of a single component plasma, there is no resonance at the individual carrier cyclotron frequencies. However, there are now other resonances, even in low fields at low frequencies. These are the so called hybrid resonances, known from gaseous plasma 's and studied in Bi for the first time by Smith, Hebel and Buchsbaum[7]. For a plasma consisting of an isotropic group of electrons and an isotropic group of holes, in the ordinary polarization the dielectric constant is still independent of field, but in the extra ordinary polarization the effective dielectric constant becomes[7]:

$$\varepsilon_{\mathrm{eff},\perp} = \varepsilon_1\left(1 - \frac{\omega_p^2(\omega^2-\omega_p^2-\omega_{ce}\omega_{ch})}{(\omega^2-\omega_{ce}^2)(\omega^2-\omega_{ch}^2) - \omega_p^2(\omega^2-\omega_{ce}\omega_{ch})}\right) \tag{7}$$

where ω_p is now defined as $\omega_p^2 = (ne^2(m_e+m_h))/m_e m_h \varepsilon_o \varepsilon_1$.
m_e and m_h are the effective masses of electrons and holes respectively and $n=n_e=n_h$ is the density of one group of carriers and the plasma is assumed to be compensated. ω_{ce} and ω_{ch} are the cyclotron frequencies of electrons and holes respectively.

Under typical microwave conditions, $\omega^2 << \omega_p^2$, the case investigated in ref. 7, it is seen from (7) that there is a resonance given very nearly by:

$$\omega = \sqrt{\omega_{ce}\omega_{ch}} \tag{8}$$

which is the so called electron-hole hybrid resonance.

It was shown[7], that at the resonant field the two counterrotating electron and hole charge clouds establish an almost purely longitudinal polarized field, resulting in small space charge oscillations at the frequency $\sqrt{\omega_{ce}\omega_{ch}}$. At the hybrid resonance, the rotation sense of the elliptically polarized, total field changes sign, such that the plasma no longer screens external fields. The propagating fields are termed Alfvèn waves in the high field limit $\omega << \omega_c$. They are typical for a compensated plasma and are the analog of Helicon waves in a single component plasma. Their classical dispersion relations follow from (7) with the assumption $\omega^2 < \omega_{ce}^2, \omega_{ch}^2, \omega_p^2$ In general, if the plasma has p groups of carriers, p-1 hybrid resonances will result[7]. This is the case for Bi and Sb. Between two resonances, the dielectric constant goes through zero and a stop window for propagation of waves is created between a dielectric anomaly and the next hybrid resonance. Expressions for the hybrid frequencies involve in general the cyclotron frequencies of the different carrier groups and their relative densities.

The situation becomes still more complicated when the carriers possess anisotropic masses[7]. In the ordinary polarization (E||B), cyclotron resonance can now occur for $\omega << \omega_p$ in cases where the plane of the electron orbit is no longer perpendicular to the magnetic field, e.g. when the magnetic field is not directed along a principal axis of an ellipsoidal mass tensor. Unlike the isotropic case, the E-vector has now a component in the orbit plane and a so called tilted orbit cyclotron resonance[7] (CR) occurs. At a field, slightly higher than the CR field, the effective dielectric constant becomes negative and field independent again, at a dielectric anomaly (DA) whose position depends on the anisotropy. In the other polarization mode (E ⊥ B) it was shown[7] that Hall effect screening is no longer completely effective when carriers are distributed over several nonequivalent mass tensors. Thus, in addition to the hybrid resonances, cyclotron resonances may be possible also.

Fig. 2 illustrates the (low-frequency) situation as it applies to Bi, for B parallel to a binary axis, the case most extensively studied in this work. With the magnetic field along a binary axis in the trigonal plane, electrons are distributed over 2 equivalent ellipsoids, whose long axes are inclined almost 60° with respect to the magnetic field (B,C pockets) and one ellipsoid which has one of the axes parallel to the field (A electrons).

Therefore, for $\vec{E}//\vec{B}$, tilted orbit cyclotron resonance of the BC electrons occurs (onset of the peak in Fig. 2 (E//B), closely followed by the DA. In fact, as can be seen in Fig. 2 (E//B), the largest change in absorption occurs at the DA, not at CR.

For E ⊥ B, two hybrid resonances occur, whose positions determine the onsets of absorption in Fig. 2 (E ⊥ B). The first is an electron-electron hybrid, the other an electron-hole hybrid resonance. In between, a dielectric anomaly occurs. The field region much above the second hybrid is the range of classical Alfvèn wave propagation.

For the bisectrix direction, two tilted orbit resonances occur for E//B (although the one due to the A electrons is very weak because the orbit is only slightly tilted for this orientation). For E ⊥ B, the calculated absorption spectrum is very similar to Fig. 2[7] (E ⊥ B).

The complete conductivity tensor appropriate for Bi was derived by Lax et al.[41] and Smith et al[7]. The explicit expressions as given in ref. 7 have been used to draw Fig. 2 and will be used in the next section. The same conductivity tensor was also used for Sb. However, unlike the Bi case, the hole pocket is now not located at the high symmetry point T in the BZ, but slightly shifted from it[42]. As a result of symmetry, there must then be either 3 or 6 pockets within the BZ, depending on whether they are shifted along the ΓT-line or away from it. It seems generally believed by now that there are 6 hole pockets[42]. As each pocket contributes additively to the total conductivity and because they are two by two identical for symmetry reasons, the two possibilities are indistinguishable within the model used. Each pocket is assumed to be ellipsoidal, although distinct deviations from ellipsoidity are known to exist[42]. Because of the more complicated Fermi surface, on the basis of the above model, tilted orbit cyclotron resonance can be expected from electron- as well as hole orbits, while there is a larger number of hybrid resonances as compared to Bi. This will be further discussed in connection with the experimental result.

3 ii. Inclusion of quantization effects.

The theory, outlined above, remarkably well describes magnetoplasma effects in Bi at microwave frequencies and at the corresponding low magnetic fields. As the magnetic fields approach the extreme quantum limit however, quantum effects must be taken into account.

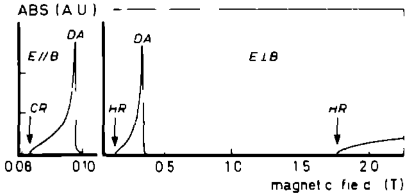
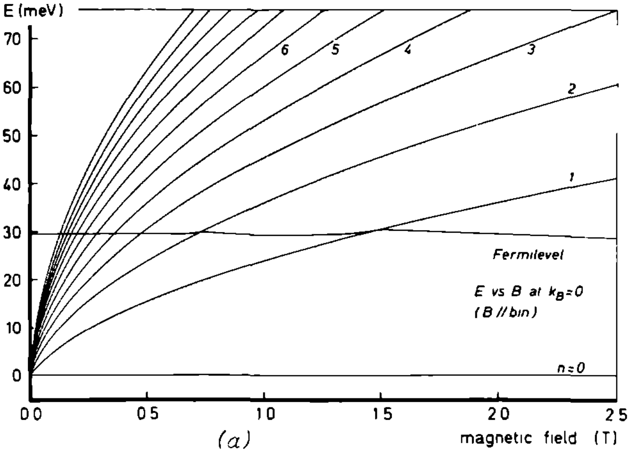
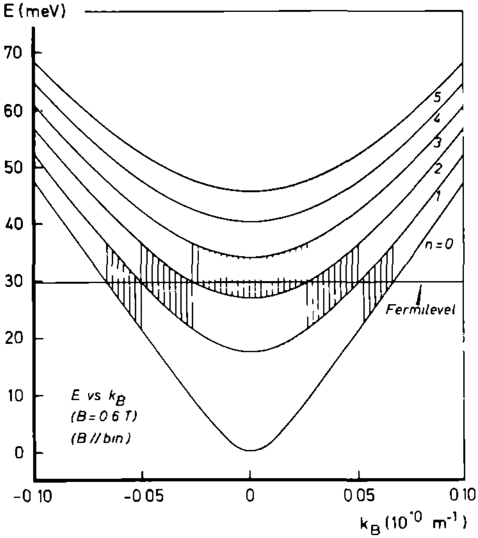


Fig. 2. Theoretical absorption behaviour of Bi as a function of magnetic field due to plasma resonances and dielectric anomalies according to the classical theory (ref. 7). The curves were drawn for ν=250 GHz and ωτ>>1. Voigt configuration, magnetic field parallel to binary axis. Left: ordinary polarization, right: extra ordinary polarization.



(a)



(b)

Fig. 3. (a): Landau level energies (n,s=-½) at k_B=0 versus magnetic field (without spin splitting). Magnetic field along binary axis. (b): k_B-dependence of Landau level energies at fixed magnetic field for (n,s=-½) levels. Magnetic field along binary axis. Vertical lines connect occupied and adjacent unoccupied states.

It is well known[2] that as a result of the small bandgap between valence and conduction bands at the L-point in the Brillouin zone, the energy dispersion of the electrons strongly deviates from parabolic. The appropriate dispersion relation is very well approximated by the two band model[2]. The energy levels in a magnetic field are then given by:

$$E_{nsk_B} = [\,(E_g/2)^2 + E_g((n+\tfrac{1}{2}-s)\hbar\omega_c + \hbar^2 k_B^2/2m_H)\,]^{\frac{1}{2}} \qquad (9)$$

Here, and in what follows, the zero of energy is taken midway between the valence and conduction bands. E_g is the energy gap, n and s are orbital and spin quantum numbers ($n = 0,1,2.....$, $s = \pm\frac{1}{2}$), k_B is the component of momentum along the magnetic field, $\omega_c = eB/m_c$ and m_c and m_H are the cyclotron and longitudinal effective mass, evaluated at the bottom of the band.
Masses at the Fermi energy, E_F, (indexed by "F") are related to those at the bottom of the band by

$$m_F = (2E_F/E_g)m \qquad (10)$$

where E_F is the Fermilevel with respect to the zero of energy defined above. In Fig. 3a the Landau level position at $k_B=0$ is plotted against magnetic field using (9). The k_B-dependence at fixed magnetic field is shown in Fig. 3b. Fig. 3 corresponds to the situation with B along a binary axis.

As expressed by (9), spin and Landau levels are degenerate in Bi within the two band model[43]. Experimentally however, a spin splitting in the order of 10% for the light electrons exists and can be described by introducing a spin mass tensor which differs from the orbital mass tensor[44]. The term $s\hbar\omega_c$ in (9) is then to be replaced by $s.\hbar.eB/m_s$ (SBR-model). The spin splitting is due to coupling between more than two bands[45]. As no spin splitting is observed in interband transitions[28], the general theoretical expression[45] could be simplified to give energy levels somewhat different from the SBR-model:

$$E_{nsk_B} = [\,(E_g/2)^2 + E_g((n+\tfrac{1}{2}-s)\hbar\omega_c + \hbar^2 k_B^2/2m_H)]^{\frac{1}{2}} + s\,G\,\hbar\omega_c \qquad (11a)$$

which applies if $n+\frac{1}{2}-s \neq 0$. G is a spin splitting parameter and

$$E_{nsk_B} = [\,(E_g/2)^2 + E_g\,\hbar^2 k_B^2/2m_H]^{\frac{1}{2}} + a\,B \qquad (11b)$$

which holds for $n+\frac{1}{2}-s=0$. a is a constant close to zero. Eq. (11) is strictly valid only at $k_B = 0$ and holds for each electron pocket with the appropriate parameters. It should be noted that with (11) the spin splitting of cyclotron resonance, expected with the SBR level scheme, is also removed (except for transitions starting from the lowest level). Equation (9)(with SBR-splitting) and (11) yield furthermore almost the same Landaulevel spectrum for energies below the Fermilevel, except for the lowest level, which is undefined in fields $\geq$ 2 T using (9) for the light electrons. Eq. (11b) holds for fields $\lesssim$ 10 T[30].

The total conductivity tensor for Bi is a summation of contributions from different pockets. Each contribution is proportional to the density of electrons in that pocket. In ref. 7, the carrier distribution over different pockets was assumed independent of magnetic field and equal to the zero field distribution. In fields near the quantum limit, carrier density is strongly dependent on field and shows oscillations corresponding to Landau levels crossing the Fermilevel . The Fermilevel and carrier density in each pocket can be calculated from the condition

$$\Sigma\, n_i = p \tag{12}$$

which must hold at all fields. n_i is the electron density in electron pocket i and p is the hole density. n_i, p are found by summing over the Landau levels below the Fermilevel using standard formulas[44]. The energy levels were assumed to depend on field according to SBR[44] and (11b)[32], if spin splitting was included. This field dependent carrier density is now phenomenologically used in the classical expression of Smith et al[7]. This procedure was used previously also in ref. 24 en 46. In addition, as the Fermilevel varies, the mass parameters will also vary, in accordance with (10).

For a more rigorous description, it is necessary to use the quantummechanical expression for the optical conductivity tensor components:

$$\sigma_{\mu\nu}(\omega,q) = -e^2 \sum_{\nu\nu'} \frac{(f(E_\nu) - f(E_{\nu'}))\ <\nu|v_\mu(q)|\nu'><\nu|v_\nu(q)|\nu'>^*}{(E_\nu - E_{\nu'})\{(i/\hbar)(E_\nu - E_{\nu'}) - i\omega\}} \tag{13}$$

which is a well known result from first order time dependent perturbation theory. In expression (13), ν' and ν stand for a complete set of quantumnumbers, $f(E_\nu)$ is the occupation factor for a state ν with energy E_ν and $<\nu|\vec{v}(q)|\nu'>$ is a velocity matrix element. q is the wavevector of the selfconsistent field in the medium, which will be retained for the moment, but will be put equal to zero finally. For an isotropic and for an ellipsoidal (parabolic) Fermi surface, with the magnetic field in the plane defined by two principal ellipsoid axes, the matrix elements were evaluated explicitly in ref. 47 and 48, using a current density operator formalism. The more general expressions, with the magnetic field in an arbitrary direction, are stated in ref. 20. However, as the Bi bandstructure is highly nonparabolic, these expressions are inadequate. Blewitt and Sievers[20] therefore introduced an effective cyclotron frequency, which depended on B in a nonlinear fashion and was some average over the Fermisurface of the Landau level energy differences consistent with (9). As our experiments are especially aimed at investigating quantum behaviour of the plasma near the extreme quantum limit, the nonparabolic band was taken into account in detail . For this purpose, the velocity matrix elements for the two band model, as calculated by Wolff,[49] have been used. They are conveniently expressed in terms of the matrix elements of the effective mass (parabolic) Hamiltonian, which are well known. The eigenfunctions ψ of the two band Hamiltonian can be expressed in eigenfunctions χ of the effective mass Hamiltonian H^*[49]:

$$H^* = (\vec{\pi}.\ \vec{\alpha}.\ \vec{\pi})/2 - \vec{\mu}.\ \vec{B} \qquad (14)$$

where $\vec{\pi} = \vec{k} - e\vec{A}/c$, the canonical momentum, $\vec{A}$ vector potential, e electron charge, c speed of light, $\vec{\alpha}$ the inverse effective mass tensor, $\vec{B}$ the magnetic field and $\vec{\mu}$ a (2x2) matrix vector, representing the effective magnetic moment (g-factor) of the electron spin.

Eigenvalues and eigenfunctions of (14) are well known:

$$\varepsilon = (n + \tfrac{1}{2} \pm \tfrac{1}{2})\hbar\omega_c + \hbar^2 k_B^2/2m_H \qquad (15)$$

for the energy eigenvalue while the orbital part of the eigen functions consists of products of plane waves and harmonic oscillator functions. Note that spin levels are degenerate with the Landau levels. The velocity matrix element of the effective mass Hamiltonian does not couple different spin states.

In the two band Hamiltonian different spin states are mixed, if spin orbit coupling is taken into account. We will however omit the mixing term, because then the matrix element becomes very simple and directly expressed in terms of the well known velocity matrix elements of the effective mass Hamiltonian[49]:

$$< \psi|\vec{v}_{2\ \mathrm{band}}|\psi'> = \sqrt{\frac{E_g}{2E_\nu}}\ \sqrt{\frac{E_g}{2E_{\nu'}}}\ \left\{ \frac{E_\nu + E_{\nu'} + E_g}{[(2E_\nu + E_g)(2E_{\nu'} + E_g)]^{\frac{1}{2}}} \right\} < \chi|\vec{\alpha}\ .\vec{\pi}\ |\chi'> \qquad (16)$$

which involves the orbital part of the wave function only. The quantum numbers n, k_z, k_y and s can still be used for the two band model. The term omitted results from spin orbit coupling and allows for the possibility of spin flip transitions, which are forbidden for a parabolic band. The relative strengths of the matrix element for spin flip and CR (16) are estimated by Wolff[49]:

$$\left| \frac{M_{\mathrm{spin}}}{M_{\mathrm{cyclotron}}} \right|^2 \sim \left\{ \frac{E_\nu - E_{\nu'}}{2\,(E_F + E_g)} \right\}^2 \qquad (17)$$

which is a clearly negligible ratio at microwave frequencies. Under typical conditions of our experiment, this ratio can be of order of 10% or even higher. Therefore, although (16) will still give the dominant contribution, it should be kept in mind that the mixing term is not necessarily negligible . The same approximation for the matrix element was also made by Miclavc et al.[50] to treat the FIR Azbel-Kaner effect in Bi.

The simplification we made now, has an easy physical interpretation. In eq. (13), for the energy eigenvalues we use the appropriate expressions as follow from the two band model (e.g. eq. (9)). For the matrix element, the results of the parabolic case[20,47,48] can be used, with mass parameters at the bottom of the band, however multiplied by a factor as given in (16). The factor in front of the right hand side of eq. (16) can be viewed as describing the energy dependence of the effective masses as they appear in the matrix element. The velocity matrix element is inversely proportional to 1/m*. The factor in curled parentheses is close to unity for low fields, while the factor $\sqrt{E_g/2E_\nu}\sqrt{E_g/2E_{\nu'}}$ gives the effective energy dependence of the mass and should be compared with (10). Although (16) is strictly valid for the two band model only, i.e. with degenerate spin and Landau levels, a spin splitting might be introduced phenomenologically by using the expressions (11) in (16). It will be discussed in the next section, that this is not essential for analyzing the experimental results.

Using a simpler Hamiltonian than Wolff[49] by a more simple spin orbit coupling, Hansen[51], following Arora et al.[52], has recently derived a different expression for the matrix element. It is of the form of (16) but the multiplication factor differs. A previous expression given by Arora was corrected. It is of interest to compare the different multiplication factors (see eq. (16)) as they appear in the literature:

Wolff[49]

$$\frac{E_g}{2} \frac{1}{\sqrt{E_\nu E_{\nu'}}} \left\{ \frac{E_\nu + E_{\nu'} + E_g}{[(2E_\nu + E_g)(2E_{\nu'} + E_g)]^{\frac{1}{2}}} \right\} \qquad (18a)$$

Arora et al.[52]

$$\frac{E_g}{2} \left[\frac{1}{2E_{\nu'}} + \frac{1}{2E_\nu} \right] \qquad (18b)$$

Hansen[51]

$$\frac{E_g}{2} \frac{1}{(E_\nu + E_{\nu'})/2} \qquad (18c)$$

Apart from the last factor in (18a), these expressions prescribe three different ways of averaging the effective mass in a matrix element describing transitions between different Landau levels. In fields below the quantum limit, the last factor of (18a) is close to unity and all the expressions yield numerically approximately the same values. At higher fields however, the numerical results are considerably different. Especially, the last factor of (18a) deviates from unity and increases with B as $\sqrt{B}$. The analysis of the experimental results will be based essentially on (18a), but the other matrix elements will be investigated also. As will be shown, the use of (18c) leads to the best agreement with the experimental results.

In the analysis as described above, all specific qualitative features of the two band model are neglected, but the nonparabolic bandstructure, which manifests itself strongly in the experiments, is taken into account. Having written the inverse effective mass tensor of the carrier pocket in question with respect to a coordinate frame in which the magnetic field is along the z-axis, the matrix elements can be found from Blewitt and Sievers[20], generalizing the results of Quinn[48]. To proceed we will work in the strictly local limit i.e. q=0. This greatly simplifies the matrix elements and thereby also the summation in eq. (13). Moreover, calculation of the electrodynamic properties (reflection, absorption) for a general $\varepsilon(q,\omega)$ is a boundary value problem of great mathematical complexity and cannot be solved generally. In neglecting the q-dependence of σ, some typical nonlocal effects as the Azbel-Kaner effect[40], which are clearly present in the experiments, are skipped in the analysis.

The overlap integral $f_{nn'}$ between harmonic oscillator functions as it appears in the matrix element reduces to: $f_{n'n} = \delta_{n'n}$ (Kronecker δ) and the elements are proportional to $\delta(k_y,k_y)\delta(k_z,k_z)\delta_{s's}$. Only states with n'=n or n ± 1 are coupled (selection rule $\Delta n=0, \pm 1$). Summation over k_y, k_z is replaced by an integral, where integration over k_y is trivial, because the summand doesn't depend on k_y so that we have the usual replacement:

$$\sum_{\substack{k_y\, k_z\\ nn's}} \rightarrow \frac{2}{(2\pi)^2}\frac{eB}{\hbar} \sum_{\substack{nn's\\ (n'=n\pm 1)\\ =n}} \int_0^{k_{max}} dk_z \qquad (19)$$

Summation is over occupied levels only, so that k_{max} for each Landau level follows from (9), with $E_{nsk_{max}} = E_F$. The Fermi level is calculated as described before. For those tensor components where the matrix element is nonzero for

$n' = n$, the factor $(f(E_\nu) - f(E_{\nu'}))/(E_{\nu'} - E_\nu)$ in (13) was replaced by $(df/dE)/_{E_F}$ which was assumed to be a δ-function.

Influence of the finite relaxation time was taken into account in the usual way replacing $\omega \rightarrow \bar{\omega} = \omega - i/\tau$ in eq. (13). A refined way of introducing a relaxation time exists[53], but this reduces also to the simple prescription for our typical condition $\hbar/\tau >> E_{\nu'} - E_\nu$. In calculating the Fermilevel, level broadening due to the finite temperature and relaxation rate has been neglected. Also, the occupation factors are assumed to be 0 or 1 when the level is above or below the Fermilevel respectively.

3 iii. Comparison with other work.

For interpreting infrared magneto-reflection data, Vecchi et al.[30] have taken into account modifications to the two band model according to Baraff[45]. Also matrix elements were evaluated including spin effects. A considerable amount of literature is available dealing with nonlocal effects in a low density plasma such as Bi. In their original work on far-infrared cyclotron resonance in Bi, Strom et al.[18,19] have approached the problem from the extreme nonlocal theory which gives the Azbel-Kaner effect in metals[40]. This theory was then modified towards the more local case in terms of retardation effects as treated by Drew [54] and Meierovich[55]. More recently, Miclavc and Drew[50] derived an approximate nonlocal theory which included effects due to wave propagation in the bulk as well as quantum effects. This theory could be used to analyze the subharmonic Azbel-Kaner resonances observed in ref. 19.

To explain the strong hole cyclotron resonances in ref. 7, which don't follow from a local theory, Hebel extended the strictly local theory[7] by taking the current density to be proportional to a power series of $E(\vec{r})$. Solving the wave equations when limiting to terms of second order gave the fundamental resonance, while higher order resonances appear in the theory by taking higher order derivatives. Nonlocal and wave propagation effects at microwave hole cyclotron resonances and subharmonics were also studied by Nakahara et al.[56]

Recently, Giura et al.[57] have observed and analyzed for the holes in Bi at 50 GHz the electromagnetic analogue of the giant oscillations in the absorption of ultrasound[58], which essentially is a nonlocal effect.

In a series of papers, Braune, Lebech and Saermark and coworkers have studied in detail propagating wave effects in Bi and BiSb-alloys at frequencies up to 300 GHz. From observations of Fabry-Perot or Rayleigh interferences, distinct deviations from classical (local) Alfvèn wave dispersion relations are observed[59-61]. Infinite medium dispersion relations were solved with numerical methods, using a nonlocal magnetoconductivity tensor[62,63] and compared with the experiments[38]. Nonlocal effects in the Alfvèn wave dispersion were also considered by Yokota[64] and studied in Bi by Lupatkin and Nanney[65] and Guthman et al.[66]

From the foregoing, it is clear that the use of a strictly local theory has questionable validity. However, all existing theories are necessarily approximate and usually applied to particular cases. A theoretical framework which describes a variety of magnetoplasma effects in a frequency and magnetic field range, which span almost two orders of magnitude, is hard to obtain. On the other hand, local classical theory explains the gross features of magnetoplasma effects very well at low frequencies[7]. At higher frequencies, and correspondingly higher fields, nonlocal effects are not expected to become more important.

Systematic studies of these classical magnetoplasma effects as hybrid resonances and dielectric anomalies at frequencies above 300 GHz have not been published. It is the purpose of the present work to investigate experimentally how the classical behaviour is modified as a result of quantization of the charge carriers. Therefore, an extension of the classical theory, to take into account the Landau level quantization, but neglecting effects due to spin and non-locality should be very well suited as a reference for comparing the experimental results. This is moreover relevant, as concepts like HR and DA start to lose their meaning from a nonlocal point of view[63]. In this respect, the clear identification of HR's and DA's in Sb, with a two orders of magnitude higher carrier density as compared to Bi, is of interest. The calculated results from the local theory however, have not much resemblance with the data in case of Sb as far as line shapes and signal magnitudes are concerned.

4. Experimental results and analysis.

4 A. Bi, B// binary axis.

In the low frequency part of the range covered, the present experiments link up with the microwave results. Some observed spectra in this range are given in Fig. 4. The radiation is unpolarized, so that effects due to the ordinary (E//B) and extra ordinary (E ⊥ B) wave are simultaneously present. The three features, depicted schematically in Fig. 2, can be identified in Fig. 4. The smallest peak at low field is associated with the light electron (B,C) tilted orbit cyclotron resonance and originates from the ordinary polarization mode. The second peak is a dielectric anomaly in the extra ordinary mode, preceded by an electron-electron hybrid resonance. The electron-hole hybrid resonance, also occurring for $E \perp B$, can be seen at 245 GHz only. This high-field hybrid resonance was not investigated further.

The absorption peaks in Fig. 4 should, according to the theory, be viewed as a window, where wave propagation occurs. The window opens at a resonance, either CR or HR and closes at a dielectric anomaly. For low $\omega\tau$-values, the window edges are not sharply defined and the window appears rather as a single peak. As the strongest change in absorption is due to the dielectric anomaly, the peak position is determined mainly by the (broadened) dielectric anomaly. The window character becomes more pronounced at higher frequencies.

4 A i. Parameters.

Before comparing the results with the calculations, it is necessary to discuss the parameters which have been chosen in the calculated curves.

A very important parameter is the relaxation time. It depends on the sample and must be adjusted in accordance with the experimental results. A constant (magnetic field and momentum independent) relaxation time τ has been used, as prescribed in the previous section . Relaxation times for electrons and holes however were in principal allowed to be different. The relaxation times were then adjusted such that the calculated absorption curves are in fair agreement with the experiments at the lower frequencies of Fig. 4. A detailed fit however cannot be obtained within the model used, by varying τ only. Thus the procedure cannot be used to derive accurate and reliable values for τ and the values obtained should be regarded as rough estimates only.

It turns out to be satisfactory to use the same value for electrons and holes, if no detailed fit on a specific part of the absorption curve is attempted. A value $\tau \sim 6.10^{-12}$ s gives satisfactory qualitative agreement of the line shapes (disregarding the peak offsets between experimental and calculated results) at low frequencies (see Fig. 4). The agreement deteriorates qualitatively, if this value is changed by more than about 50%. If only part of the curve is fitted, for instance the slope of one of the dielectric anomalies, as done in other work, the fit procedure is sensitive for the value of τ as well as on the ratio of relaxation times for electrons and holes. There is however strong indication that unique relaxation time parameters, fitting more than one lineshape or slope, cannot be found.

It is well known that the relaxation time strongly depends on frequency in this frequency range , although the exact frequency dependence is not very well established. CR experiments at very low frequency suggest a frequency dependence as ω^2 or faster[67]. By fitting the slope of the low field DA such as given in Fig. 2, at a laser frequency of 891 GHz, Drew et al.[23] obtained a value for the relaxation rate much lower than expected from the results of ref. 67, but they suggest a ω^2-dependence also. From broad band measurements at large magnetic fields, Blewitt and Sievers[20] derived a power law dependence for the relaxation rate of approximately $\omega^{3/2}$. From a fit of the slope of the low field dielectric anomaly, Kamgar et al.[24] found a linear frequency dependence for τ with B along a bisectrix axis, but no well defined frequency dependence with B along a binary direction. From this and other work, it is shown that the relaxation rate is anisotropic, as well as magnetic field dependent. A magnetic field dependence of the relaxation time was observed in fields below the quantum limit from an analysis of Alfvèn wave interference patterns, which eliminated possible charge carrier oscillations[68]. The strongly increased transmission of FIR radiation at fields where low order electron Landau levels pass the Fermilevel is also attributed to a change in relaxation time[20].

Electron-phonon scattering or electron-electron scattering are thought to be responsible for the frequency-dependence of the scattering rate, but no unambiguous interpretation exists as yet[24].

The value of the relaxation time at the lowest frequencies used in the present work is considerably smaller than the frequency- or temperature limited value as measured by Drew et al.[23] and Kamgar et al.[24] At higher frequencies however, the frequency-dependent term is nevertheless expected to dominate and therefore a frequency-dependence of τ should be included in the calculations. The observed residual relaxation time ($\sim 6.10^{-12}$ s) is somewhat lower than that found from comparable experiments ($2.3.10^{-11}$ s[7], 10^{-11}-10^{-10}s [24]) and much shorter than values reported for very carefully prepared and zone-refined samples. As mentioned before, some strain in the sample might partly account for the rather low relaxation time. It should be noted on the other hand that the crystals were used as grown, but were stored at room temperature for several years, which seems to be an excellent annealing condition [4].

From the foregoing, it is clear that there is no conclusive way how to include scattering in the calculations. As a reasonable guide, the following expression was used for τ (assuming the same values for electrons and holes):

$$\tau^{-1} = \tau_o^{-1} + A\,\omega^{3/2} \tag{20}$$

where τ_o is the residual relaxation time. A was chosen such that the frequency dependent term gave the result of Drew et al.[23] at 891 GHz and $\tau_o \simeq 6.10^{-12}$ s. For the calculation, no broadening of the Landau levels due to finite temperature and relaxation time was included. This means that Shubnikov-deHaas related effects will not be properly smeared in the calculation.

The next important set of input parameters for the calculations are those describing the band structure. The pockets of Fermi surface are assumed to be ellipsoidal and are described by an effective mass tensor. With the usual convention for labeling the binary, bisectrix and trigonal axes as 1,2 and 3 the principal ellipsoid for the electrons (A-electrons) has the form (21), where m_o is the free electron mass.

$$\overset{\rightarrow}{\vec{m}}_{AF} = \begin{pmatrix} m_{1F} & 0 & 0 \\ 0 & m_{2F} & m_{4F} \\ 0 & m_{4F} & m_{3F} \end{pmatrix} m_o \tag{21}$$

Mass tensors for the B, C-electrons are obtained form (21) by rotating the coordinate frame over ± 120° with respect to the trigonal axis. Longitudinal and cyclotron masses follow from

$$\begin{aligned} m_H &= \vec{h} \cdot \overset{\rightarrow}{\vec{m}} \cdot \vec{h} \\ m_c &= \{\det \overset{\rightarrow}{\vec{m}}/m_H\}^{\frac{1}{2}} \end{aligned} \tag{22}$$

with $\vec{h}$ a unit vector in the direction of magnetic field. A spin mass tensor according to the SBR scheme[44] can be analogously defined. The off-diagonal component of $\overset{\rightarrow}{\vec{m}}_{AF}$ is a result of the slight (6.5°) tilt of the electron ellipsoid out of the trigonal plane. The hole mass tensor for Bi is diagonal, with $m_1 = m_2$. For Sb, both electron and hole mass tensors are diagonal, with the tilt larger than for Bi[42]. For Bi, the mass tensor components were obtained from the cyclotron masses along principal directions and the orientation of the ellipsoids, given by Edelman[1].

For the energy gap at the L-point, as it appears in the two band model (e.g. (9) and (10)), a value of 13.5 meV was used. Literature values for E_g vary from 11 meV[28] to 15.3 meV[44], while the average is generally accepted

(13.8 meV[29], 13±2 meV[19]) . Fermi energy values (with respect to the bottom of the electron energy band) range from 20 ± 3 meV[33], 25.4 meV[32], 27.6 meV[44], 29.8 ± 0.6[19] to 32 ± 2 meV[69]. A value of 29.7 meV has been chosen in the analysis, as this value was obtained in comparable experiments[19]. The Fermi energy is obtained from the band parameters and has in fact been fixed (at 0 field) by a proper adjustment of the band overlap energy E_o. A value of E_o = 41.1 meV is then consistent with the previous parameters (and equal to the value used in ref. 24). For the electron density, a value of $2.9 \cdot 10^{17}$ cm^{-3} is then found, consistent with literature values, which range from $2\text{-}4 \cdot 10^{17}$ cm^{-3} [1]. An anisotropic lattice dielectric constant is employed , using $\varepsilon_1 = \varepsilon_2 = 100$ [17] and $\varepsilon_3 = 60$ [32]. Literature values for $\varepsilon_{1,2}$ scatter by about 10% and those for ε_3 by about 20% [1]. In the final calculations, the strict two band model (eq. (9)) was used, but to investigate the importance of spin splitting, calculations were made also according to the SBR level scheme with spin mass parameters from ref. 44 and with the simplified Baraff model (eq. (11)). The parameters G and a in eq. (11) were taken from ref. 30 and 32 respectively. The parameters used for Bi are summarized in table I.

4 A ii. General discussion, low frequency, low field.

In Fig. 4, some results of the calculations are also plotted. The calculated curves are an average of the results obtained for the two polarization components. Three calculated curves are given. For simplicity, here and in what follows, we will designate the three types of calculation as I, II and III. This designation refers to the three models outlined in the previous section. The calculation for type I is the same as used by Smith et al.[7] and is therefore purely classical. For II, the classical method I is extended by introducing phenomenologically a magnetic field dependent carrier distribution and total density using eq. (12) and allowing the Fermi level to vary. The effective masses at the Fermi level vary according to (10). For method III, the quantummechanical expression for the optical conductivity (eq.(13)) is used, together with the matrix element (16). Thus the carrier density quantization, as well as the nonparabolic bandstructure is fully taken into account. However, only the contribution of the light electrons (BC) is calculated quantummechanically, while the contribution of the A electrons and holes is still described using the classical formulas. This is justified as the latter two carrier groups are essentially below the quantum limit (B_{qu}) in the field range considered. The cyclotron mass of the A-electrons is about ten times

$m_{1F} = 0{,}00603\ m_o$

$m_{2F} = 1{,}27$

$m_{3F} = 0{,}0268$

$m_{4F} = -0.141$

$M_1 = M_2 = 0{,}0639\ m_o$

$M_3 = 0{,}703$

$E_g = 13.5$ meV

E_F (B=0) = 29,7 meV

Eov = 41.05

$\varepsilon_1 = \varepsilon_2 = 100$

$\varepsilon_3 = 60$

$n = 2.9 \,.\, 10^{17}\ \mathrm{cm}^{-3}$

$G = 7.6 \,.\, 10^{-3}$ (binary axis, light electrons)

$7.3 \,.\, 10^{-3}$ (binary axis, light electrons)

$8.6 \,.\, 10^{-3}$ (bisectrix axis, heavy electrons)

$a = -\ 0.033$ meV/kG (binary axis, light electrons)

$-\ 0.038$ meV/kG (bisectrix axis, heavy electrons)

$= 1.9 \cdot (-0.038)$ meV/kG (bisectrix axis, light electrons)

m_{iF} electron mass tensor parameters at Fermi level.

M_i hole " " "

E_g energy gap at L-point

E_F Fermi energy at zero field (with respect to bottom band)

Eov band overlap energy

ε_i lattice dielectric constant components.

n electron density

G, a spin splitting parameters

Table I. Values for the parameters used in the theoretical models for Bi.

higher than that of the BC-electrons. The hole cyclotron mass for B along a binary direction is again about ten times higher than that of the A-electrons. Using the two band model for the A-electrons, they reach the extreme quantum limit (only one A-electron Landau level below the (lowered) Fermi energy) at about 10 T.

It is easily shown from eq. (13) and the explicit expression for the matrix element that the quantummechanical expression reduces identically to the classical one for a parabolic band and local approximation (see also ref. 20), while it is very well approximated by the classical expression for fields not very close to B_{qu}. Moreover, the approximation of using classical expressions for the A-electrons and holes is also justified by the results of the calculations for the B, C-electrons itself, as will be shown.

It can be seen from Fig. 4 that all three ways of calculation are essentially similar at low frequencies and fields, as expected. The measured ratio between the height of the small and large peak is approximately equal to the calculated ratio for all frequencies, which justifies the simple averaging procedure over both polarizations. The relative magnitude of the peaks also is determined by the ratio of hole and electron relaxation times. These times must be chosen roughly equal to give agreement with the line shape of a single peak. The observed ratio of the measured magnitude of the peak is then again consistent with a single relaxation time for electrons and holes. As the absolute value of the absorption was not measured, we have simply scaled the calculated curves to have the same peak height as the corresponding experimental ones.

For the experimental results, peak heights were measured with respect to the background level at $B = 0$. Unlike the results of a local calculation, experimentally a strong absorption is always present, also at $B = 0$. This can evidently not be explained by adjusting the relaxation rate, but is a result of nonlocal effects. In the frequency range of Fig. 4, the measured zero field absorption is in the order of 50% of the ordinary peak and decreases rapidly with increasing frequency with respect to this peak. The zero field absorption is also strongly present in microwave experiments[7]. This (nonlocal) background absorption shows structure which is related to the Azbel-Kaner (AK) effect. A small, stepwise increase in absorption occurs at the second harmonic of the fundamental light electron cyclotron resonance fields, in Fig. 4 only visible at 1017 GHz. The skindepth δ_0 at zero field is in the order of a few μm. The cyclotron radius of the light electrons, as estimated from $R_c \sim v_F/\omega$ (v_F Fermi-velocity) (or quantummechanically $R_c \sim (2(n+\frac{1}{2})\hbar/eB)^{\frac{1}{2}}$), is $R_c \sim 0.05$ μm at 1T.

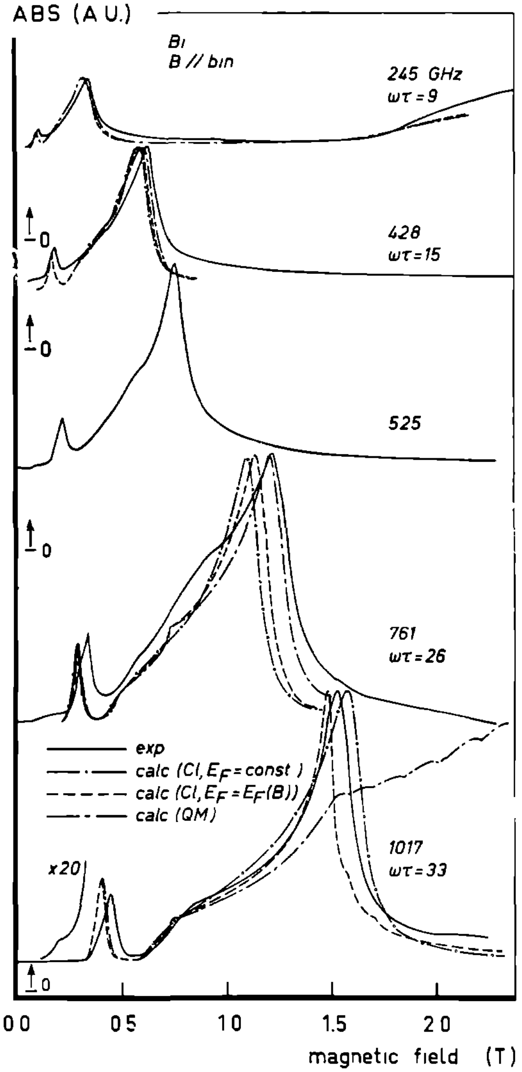


Fig. 4. Absorption as a function of magnetic field in the low frequency range. Experimental results and calculations. ωτ-values used for the calculations are indicated. At some frequencies, the zero absorption level is also indicated. Calculated curves are averages over both polarizations.

Therefore we have $R_c << \delta_o$ under our typical conditions and the skin effect is to a great extent normal. (In a magnetic field, the relevant length is R_c rather than the mean free path l; as necessarily $\omega\tau > 1, \omega_c\tau > 1$, we have $l > R_c$). Clearly the Azbel-Kaner treatment, which assumes an extremely anomalous skin effect $R_c >> \delta_o$ is not very useful here. Although a line shape analysis of subharmonic resonances will not be attempted , it is seen that the AK-formula does not apply. Use of the AK-theory should be reasonable as far as the Fermiology is concerned, because all electrons have the same cyclotron frequency at the Fermisurface (this is no longer true in high fields for a nonparabolic band). The AK-expression shows an oscillatory behaviour of the surface impedance which is periodic in ω/ω_c and decays only weakly with harmonic number. In Bi we observe only structure at the second subharmonic, which appears as a step in the absorption and is negligibly small compared to the structure at the fundamental. Some care is however required in analyzing the subharmonic line shape because it might be influenced by the foot of the fundamental (local) cyclotron resonance. Moreover, subharmonics can appear for the ordinary[19] as well as the extra ordinary[18] polarization, and these might be mixed in the present experiments.

These cyclotron resonances were analyzed in ref. 18 and 19, where also even higher harmonics could be observed at FIR-frequencies, using a field modulation technique. Miclavc and Drew[50] have recently analyzed these subharmonic lineshapes in rather detail. To calculate the surface impedance from a nonlocal theory, two different aspects naturally show up. First, there is the rapid decay of the EM-field from the metal-vacuum interface into the metal ("surface", skin effect). This accounts qualitatively for the AK-effect. Next, the surface impedance is affected by fields which at or near a resonance penetrate into the bulk to a distance which is much larger than the skin depth. The field distribution inside the metal must now be calculated using eq. (1), however with the q-dependence of σ retained and the dispersion relation $q(\omega,B)$ must be calculated selfconsistently. The resulting waves are termed cyclotron waves. The condition for nonlocal wave behaviour in the bulk is usually written as $qR_c >> 1$ where R_c is the cyclotron radius and q the wavevector of the wave. It was found in ref. 50 that cyclotron waves strongly affect the high-field side of the subharmonic resonances. The wavelength of the (local) Alfvèn waves at $B \sim 2$ T and $\nu \sim 300$ GHz is in the order of 15 µm, which is much larger than the light electron cyclotron radius (~ 0.05 µm). This indicates that local theory should still be very reasonable as a first approximation, especially when considering absorption envelopes only, without obtaining the dispersion relation. At a resonance however, cyclotron or hybrid, the dielectric constant becomes large and therefore the wavelength tends to zero for large relaxation times. Each resonance obtained in a local theory is in fact inconsistent with the assumption of small q. Nonlocal effects are therefore expected to be strongest near a resonance.

In fact, a type of wave propagation in the field range near the electron-hole hybrid resonance (high-field hybrid in Fig. 4 at 245 GHz) has recently been observed at 300 GHz[61]. These waves could not be explained using a local theory, but their occurrence seems also related to the cyclotron resonance for the A-electrons (1.04 T at 245 GHz) and the fundamental hole-CR which is close to the electron-hole hybrid. Both the A-electron- and hole-CR do not occur in a local theory.

With the relaxation time adjusted to the low field peaks, the structure at the electron-hole hybrid in Fig. 4 (245 GHz) appears somewhat sharper than in the calculated curves. This might be due to a magnetic field dependence of τ, but is likely also connected to nonlocal effects and the occurrence of the hole cyclotron resonance in view of the foregoing. In a nonlocal theory, the hybrid becomes less sharply defined and is smeared over an extended range of magnetic field[70]. Direct observation of wave propagation by Fabry-Perot interference effects is excluded in our measurements as the sample planes are not plane parallel. The rather high relaxation rate would presumably also strongly suppress their amplitudes with the thickness of the sample used. In the field region above the electron-hole hybrid, Alfvèn waves have been studied up to frequencies of 300 GHz[60]. The dispersion relation is strongly affected by carrier density quantum oscillations,but also by nonlocal effects.

The strongest absorption structure in Fig. 4, is the absorption starting at the electron-electron hybrid resonance and peaking at a dielectric anomaly. At the lowest frequencies of Fig. 4, we find that the peak is well reproduced by the calculations (the three models being still almost equivalent), apart from a slight discrepancy in peak position. At 300 GHz, the same peak was also observed by Braune et al[71]. They noted however that the line shape differed rather strongly from the theory[7]. They suggested that a nonlocal theory is necessary to account for their experimental lineshape, obtained from a sample with presumably a somewhat lower relaxation rate than used for the present work . This suggestion is supported by the observation of cyclotron waves[12] in the field range belonging to this absorption peak.

It is seen from Fig. 4, that at higher frequencies, the line shape of the second absorption peak becomes strongly-frequency dependent (compare, e.g. line shapes at 761 and 1017 GHz). Comparing with the calculation of model II and III, this should be attributed to the carrier density quantization. The knee in the 761 GHz data at ∿ 0.9 T would then correspond to the knee in the calculated result, which appears however at ∿ 0.75 T and corresponds to the field where the 2^- Landau level passes the Fermi level (see Fig. 3a). For simplicity, we will term this occurrence "Landau level resonance" (LL-resonance). The distinct bent in the experimental curve at 1017 GHz, located closely after the onset of the absorption, has a similar origin. In the experimental curves, the LL-resonances are strongly broadened as compared to the calculated ones because of the finite width of the levels, not taken into account in the calculations.

Some care must be taken in comparing experimental and calculated lineshapes, because there is some arbitrariness in scaling the curves. In Fig. 4, the calculated results were scaled to the magnitude of the large peak with respect to the zero field absorption. The absorption minimum between first and second peak is experimentally generally higher than calculated. Nonlocal effects might be responsible for the excess absorption, but with the finite relaxation times it is also dependent on the location of the first peak with respect to the onset of the second. There is a discrepancy between theory and calculation in the location of the first peak, as will be discussed. This scaling problem is especially felt at 761 GHz and exaggerates the discrepancy between experiment and calculation.

The experimentally identified field positions corresponding to a LL-resonance are always too high. This is consistently observed at several frequencies, where structure due to a LL-resonance was seen. This is not likely to be due to a misorientation. The cyclotron masses vary only slightly near the binary direction[4]. One of the two BC-masses becomes slightly lighter, going to the bisectrix direction. The other has become a factor of two higher when the field is along the bisectrix direction. As the light electrons contribute most effectively to the (optical) conductivity, an accidental misorientation effect would rather shift the LL-resonance positions to lower field values. For a tilt out of the trigonal plane, near a binary axis, the light mass is practically stationary[4]. Furthermore, the discrepancy was also observed after mounting and remounting the samples. A possible explanation could be the broadening of the levels in such a way that a level is effectively occupied as long as it is within $\hbar/\tau$ above the Fermi level. With a $\tau \sim 6.10^{-12}$ s, this amounts to ∿ 0.8 meV. The required shift of the Fermi level in Fig. 3a, which brings the LL-resonances at the apparently observed positions, is about 4 meV.

The possibility of an inadvertent doping effect, shifting the FL and lifting the compensation, can also be ruled out. This would change the relative densities in the carrier pockets and the masses at the FL and therefore strongly affect the positions of the dielectric anomalies also. The discrepancy is therefore not understood. It should however be noted that no comparison is made with line shape calculations taking the level broadening properly into account. For the further discussion, we will take the discrepancy for granted.

The low field, CR-related peak in Fig. 4 is essentially classical. Details of this structure will be discussed at the higher frequencies. The experimental peak position does not coincide with the calculated ones. This fact, which was noted also earlier[24], is of considerable importance for the width of the peak, and therefore experimental and calculated peak widths should not be matched by adjusting the relaxation time. It is rather the high-field slope of the small peak which most reliably reflects the relaxation time[23,24]. It appears that the low-field peak can be well accounted for by the local theory, which implies that it originates from the ordinary polarization mode. In this respect, it is of interest to note that for the same configuration applying to Fig. 4, a CR-related peak was observed of comparable strength, however for the extraordinary polarization at 300 GHz[71]. For the $E \perp B$ polarization, cyclotron resonance is in the local theory completely screened by the depolarizing fields[7], so that its occurrence was attributed completely to nonlocal effects[71].

Summarizing the discussion so far, Fig. 4 shows essentially the typical low frequency, classical results[7]. The frequencies are considerably smaller than the plasma frequency ν_p ($\nu_p \sim 4700$ GHz) while the relevant absorption structures occur in the field region below the quantum limit, B_{qu}. The experimental results are reasonably well described by the classical theory of Smith et al[7]. As the frequencies increase, the magnetic field dependence of the carrier density and corresponding redistribution of carriers over the different pockets, strongly influence the line shape. For detailed line shape analyses, this effect should certainly be taken into account, especially for high $\omega\tau$-values, even at relatively low frequencies. Direct evidence for nonlocal effects is only given by the high background absorption and the subharmonic cyclotron resonance.

4 A iii. El-El hybrid and DA at high fields, high frequencies.

We now turn attention to the second absorption structure of Fig. 4, at higher frequencies and fields. In Fig. 5, the calculated behaviour of the peak is displayed, when its position is moved in the extreme quantum limit. For these calculations and those to follow, the same relaxation times have been assumed as for the curves in the experimental data and the calculated results are also an average over the two polarization modes. At very low frequencies, the three methods of calculation give identically the same results. When the peak enters the quantum limit (at ∿ 900 GHz in Fig. 5) the quantummechanically (QM) calculated curve, III starts to deviate strongly from the other two. The field value B_{qu} is indicated by an arrow in each graph. Note that the horizontal field scale is not the same for all graphs. In the most simple model, I, which has no quantum effects included, the curve changes qualitatively nevertheless, at a frequency around 1200 GHz. The reason is that the cyclotron frequency becomes comparable to the plasmafrequency and has the effect of moving the DA to higher field. The transparent window between the electron-electron hybrid and DA thus becomes relatively larger in this frequency region. The line shapes as calculated from model II strongly reflect the Landau level resonances, especially at the field value B_{qu} where the last but one (1^-) Landau level passes the Fermi level. At high frequencies (e.g. 1839 GHz in Fig. 5), at fields above B_{qu}, the absorption increases less rapidly with field as compared to I, due to the magnetic field dependent increase in carrier density. Peak position of the DA is also somewhat different in models I and II.

For the calculations of type III (QM), the DA anomalously broadens near B_{qu}. The broadening evidently is due to the fact that the B, C-electrons do not have a sharply defined cyclotron frequency. They can be thought of as distributed over infinitely many groups, each having a cyclotron frequency equal to the spacing of Landau levels adjacent to the Fermi level, at a given wavevector below the Fermi wavevector. The distribution of frequencies which contribute to the conductivity according to eq. (13) is schematically indicated in Fig. 3b by vertical lines. The average cyclotron frequency changes rather strongly near B_{qu} and for $B > B_{qu}$, the increase is slower than linear (see also ref. 20). The result is that the peak shifts to higher fields as compared to II, the absorption is much lower than for calculations I and II, and the line shape is qualitatively different from I or II (see e.g. the calculation at 1839 GHz in Fig. 5).

At still higher frequencies ($\gtrsim$ 1839 GHz), the difference in peak position between II and III becomes gradually smaller, however the line shape remains drastically different. It should be noted, that at high fields the A-electrons

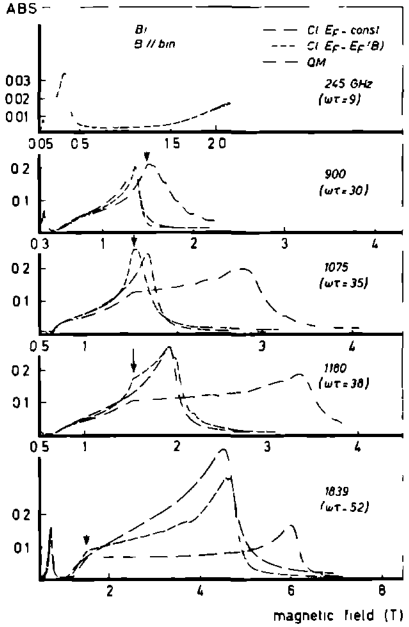


Fig. 5. Calculated absorption structure as a function of magnetic field. The behaviour of the DA peak resulting from the extra-ordinary polarization is shown when its position is moved into the quantum limit. The quantum limit field value is indicated by arrows. Magnetic field along binary axis. Realistic ωτ-values are used (see text).

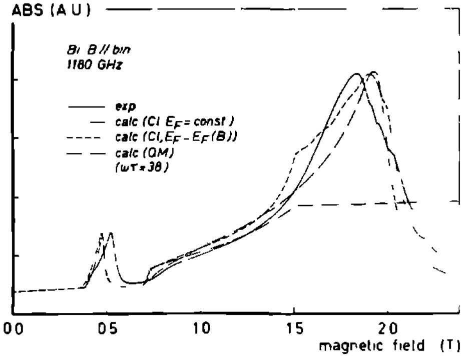


Fig. 6a. Absorption as a function of magnetic field at a frequency where a DA occurs close to the quantum limit field value. Experimental results and calculations.

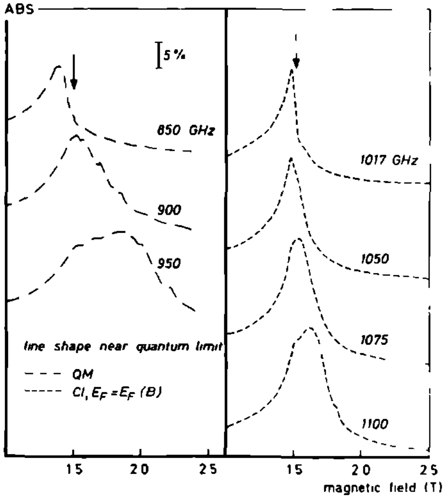


Fig. 6b. Calculated line shape evolution for a DA occurring near the quantum limit field value (B_{qu}) within the two models used. The arrows indicate B_{qu}. Realistic ωτ-values are used (see text).

also are close to the quantum limit ($B_{qu,A\ electrons} \sim 10T$) and therefore the quantum expression for the conductivity contribution of the A-electrons should be used also. From the foregoing (see also Figs. 4 and 5), it could be seen that the quantum expression however is not far from the classical one, until the very quantum limit is reached.

The above qualitative discussion of the behaviour of curves III in Fig. 5 was based on the band parameters of the two band model only, Fig. 3. As will be shown later, but should be already emphasized here, the matrix element (16) plays also a crucial role in determining the line shape and peak position of the DA.

Returning to the experimental results, at the highest frequency in Fig. 4 (1017 GHz) the curve calculated according to III strongly deviates from I and II. The experimental result however, closely follows the classical curves. The experimental peak position coincides approximately with the expected quantum limit field value B_{qu} according to Fig. 3a (1.5T). In view of the already noted discrepancy in LL-resonance position, this result might not yet be very conclusive.

In Fig. 6a, the experimental result at a somewhat higher frequency (1180 GHz) is given. The peak position now is certainly moved in the quantum limit field range, but is still very close to B_{qu}. As compared to the curves at lower frequencies, the DA seems anomalously broadened. At the high field side, an oscillatory behaviour is observed. The strong LL-resonance seen in the calculated curve, model II, cannot be clearly identified in the experimental result, but should be expected at a somewhat higher field, closer to the top, in order to be consistent with the observations at lower frequencies. A slight bending over of the rising slope, just before the top is reached, can indeed be seen, with some care. The high field slope is clearly less steep than expected, both from I and II, although the relaxation time used was proven to be very satisfactory at the lower frequencies. The result calculated according to model III, strongly differs from the calculations I and II, as well as from the observed behaviour.

To illustrate the theoretical behaviour more clearly, in Fig. 6b is plotted how line shapes evolve, when the DA position enters the quantum limit, indicated by an arrow, both for II and III. Hereby we have simply adjusted the frequencies in both cases, irrespective of the experimental frequency of Fig. 6a. The frequency differs then from the experimental one (1180 GHz) by about 30% in case III and 10% in case II. For a qualitative discussion, it is more instructive to adjust the external frequency, than to vary model para-

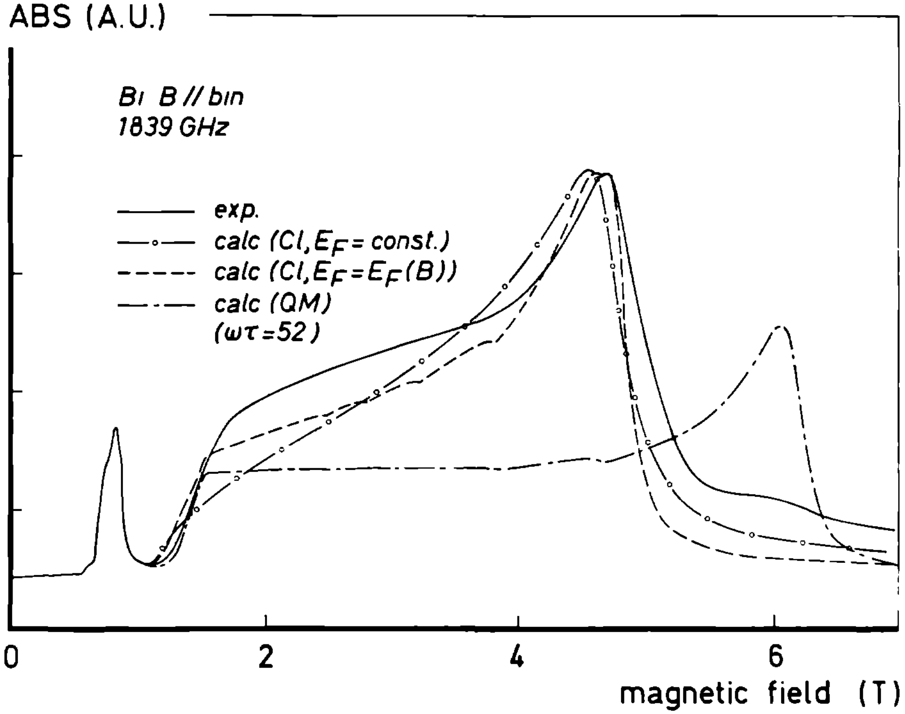


Fig. 7. Absorption as a function of magnetic field at a frequency where a DA has been driven far into the quantum limit. Experimental results and calculations.

meters. The dramatic frequency dependence of the line shape for case III is evident in Fig. 6b. At B_{qu}, the peak becomes very asymmetric, while at a somewhat higher frequency, there actually is no well defined peak at all. An oscillatory behaviour for $B > B_{qu}$ is clearly seen in the calculation and corresponds to hole Landau levels which pass through the Fermi level (causing very small FL oscillations). These oscillations would thus be properly termed hole Landau level resonances in our previous terminology. The position of the DA strongly depends on the FL and relative carrier distribution. At B_{qu}, the FL changes rapidly, but changes are fine modulated by hole LL-resonances. Thus the LL-resonances manifest themselves rather strongly in the calculation.

In the classical model II, the DA behaves more continuously as compared to III. Just before reaching B_{qu}, the peak narrows and sharpens. After entering the quantum limit, it also broadens somewhat, and a structure at B_{qu} gradually builds up. The high field slope is reasonably steep in all cases, although it is very frequency dependent also here. Hole quantum oscillations are somewhat less pronounced than in model III, but are clearly present. Comparing now the experimental result (Fig. 6a) with the calculations of Fig. 6b, it is tempting to correlate some of the observed features with qualitative aspects of model III. Especially the broadening of the experimental curve, as well as the pronounced hole quantum oscillations are striking. However, the large discrepancy in peak position remains still to be explained.

The experimental results at a frequency of 1839 GHz, where the second DA has been driven far into the quantum limit, are given in Fig. 7, together with some calculated results. The experimental curve shows a distinct bent at a field value around 1.8T. On comparison with the calculated curves (II or III) the bent corresponds evidently to the field value B_{qu}, where the $\bar{1}$ electron Landau level passes the Fermi level. Again, the experimental field value is too high. The line shape as well as the position of the DA is in surprisingly good agreement with the classical calculation and they are clearly at variance with the quantummechanical calculation. The experimental curve shows considerable absorption beyond the DA, exhibiting even some structure around 6T. Such an absorption is not accounted for by any of the calculated results I, II or III and is therefore unexplained within the model used. The coincidence of this structure with the position of the DA in model III (see Fig. 7) suggests however a possible relation. A more detailed investigation of the structure would be required for a confirmation of such a relation.

It is of importance to investigate whether the profound discrepancy between experimental results and the quantummechanical calculations, as observed in the results up to now, can be reduced by changing the model parameters. For this purpose we have phenomenologically introduced the well known deviations from the two band model, eq. (11) in the expressions for the energy values in eq. (13) and (16). As expected, the calculated results remain qualitatively unchanged. The spin splitting of the light electrons ($\sim$ 10%) appears rather pronounced at the Landau level resonances. The quantum limit field value B_{qu} is then also apparently spin splitted. This has indeed some influence on theoretical lineshapes, especially such as those of Fig. 6b. These effects have nevertheless been removed from the final calculations because they tend to complicate the calculated curves and obscure the essential features, without having a substantial contribution. Moreover, spin splitting effects are not seen in the experimental results, presumably because they are smeared out due to level broadening. The parameter a in eq. (11b) describes the field dependence of the lowest level and clearly influences the position of the DA for $B > B_{qu}$. At the high frequencies where the DA occurs in the quantum limit, the values used for the lattice dielectric constant become important. Several of the above parameters are not very accurately known, while there is also an uncertainty in the exact values to be used for the band parameters, subject to some constraints such as the effective masses at the Fermi level (see discussion on parameters at the beginning of this section). Calculations analogous to those in Fig. 5 have been carried out while varying the different parameters mentioned. It turns out that the calculated peak positions of the DA's at high frequencies can indeed be varied by as much as 10-20% for all three models, but especially model III, while keeping the parameters within reasonable limits. The large difference between model III and the experiments (Fig. 5-7) cannot be removed unless quite unrealistic parameters are chosen, inconsistent with generally accepted values. Therefore, parameter fits have not been attempted and the most simple two band model (eq. (9)) was used in the final calculation of model III. It should be remarked already here that the neglect of spin splitting does neither have severe consequences for the (spin conserving) cyclotron resonances, because spin up or spin down transitions are degenerate according to eq. (11) just as in interband transitions. Spin flip and combined resonances are neglected a priori by the choice of the matrix element (16). It will be left to a later discussion to show that the discrepancy can be largely removed by a different choice for the matrix element. The expression (18c) serves this purpose very well.

A series of results at very high frequencies (but still below the plasma frequency) is shown in Fig. 8a. The DA, which follows the electron-electron hybrid has now moved outside the field range available for the present experiments. The point to be noted here is the apparent gradual disappearance of the hybrid resonance (onset of absorption starting beyond the first peak). The position of the el-el hybrid resonance depends on the cyclotron frequencies as well as on the relative densities of both groups of carriers. In the quantum limit, the density of states of the light electron pocket becomes large, so that the electrons predominantly occupy the lowest level of the light pockets. The redistribution of carriers now has severe consequences for the hybrid resonance. Fig. 8b shows the theoretically expected behaviour. In model I, of course no irregular behaviour occurs and these curves are therefore given at one frequency only. Model II as well as model III show a qualitative similar behaviour over the frequency range of Fig. 8b. The hybrid is sharply defined at the lower frequency end, then it flattens and gradually disappears, while it becomes distinct again in the higher frequency range of Fig. 8b. Between the low-and high-frequency positions, a rather discontinuous jump has taken place. The frequency range where the transition from low to high frequency behaviour takes place however, is different in model II and III. The gradual flattening and disappearance of the hybrid is nicely seen in the experimental results (Fig. 8a). The distinct reappearance at high frequencies falls however at the end or outside the available field range.

Both model II and III behave qualitatively similar so that, on comparison with the experimental results, a striking preference for either one of the two models is difficult to make. Nevertheless, Fig. 8a shows that model II gives a very satisfactory explanation of the observed behaviour, which is clearly not improved by model III.

The carrier density distribution is very important in determining the presence or position of the hybrid and therefore, in the frequency region where the hybrid is changing character, the calculated absorption curves show rather strong hole quantum oscillations ("LL-resonances"). These oscillations are also clearly visible in the two lower experimental curves of Fig. 8a, however their amplitude is smaller than theoretically expected. A more quantitative comparison of the amplitude of these oscillations cannot be made unless the finite width of the levels is incorporated in the calculations. The occurrence of the quantum oscillations and the flattening of the absorption onset prevent a distinct designation of the hybrid resonance in the transition region (Fig. 8b).

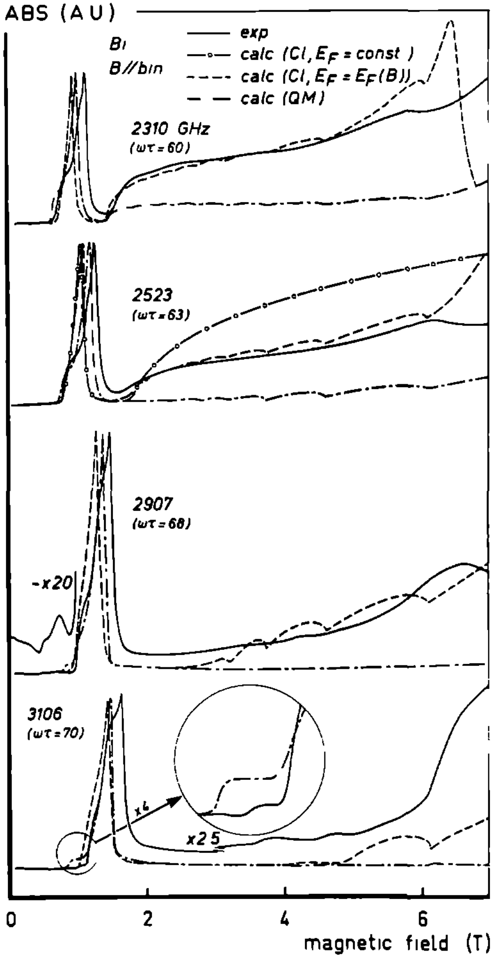


Fig. 8a. Absorption as a function of magnetic field at high frequencies, but below the plasma frequency. The position of the hybrid resonance is moved into the quantum limit. Experimental results and calculations. Calculations in the simplest model vary smoothly and continuously with frequency and are shown for one frequency only (2523 GHz).

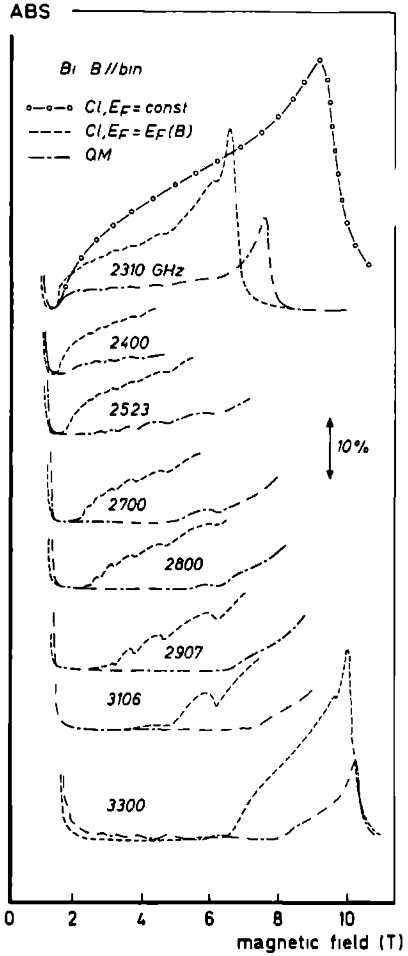


Fig. 8b. Calculated absorption line shape evolution near the hybrid resonance in a frequency range where its position occurs in the quantum limit. Realistic ωτ-values are used (see text).

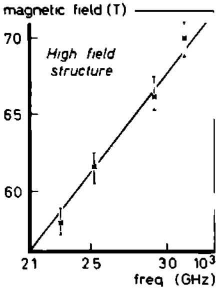


Fig. 8c. Position of the high field structure of Fig. 8a as a function of frequency.

An interesting feature of the experimental results is the high level of absorption between the first peak and the (flattened) hybrid, as compared to the absorption at zero field. The calculations predict an equal absorption level (~ 0) in both regions. This excess absorption prevents an accurate comparison of experimental and theoretical results, especially with the chosen method of scaling. It is not clear whether this excess absorption should be attributed to nonlocal effects at such high frequencies and fields. Detailed calculations of nonlocal wave propagation for the specific case of Bi have recently been carried out[63]. The dispersion equation does have many short wavelength solutions in the field range considered. Dispersion relations were calculated for frequencies an order of magnitude lower than the present ones by numerical methods. It was found[63] that the results strongly depend on frequency, but also on the relaxation time and so on a specific sample. A comparison of these calculations[63] with present experiments should therefore not be undertaken. Moreover, from a knowledge of the dispersion relation, it is still a considerable task to calculate their consequences for the surface impedance or absorption.

In the high field range of Fig. 8a ($\gtrsim 6$ T) a broad structure shows up, which is not reproduced by our calculation . The DA at these frequencies is expected to occur at, or outside, the maximum field range presently available. For 2310 GHz, it is expected at 6.5T according to II and at fields > 7 T for the higher frequencies (see Fig. 8a). It is seen at 2310 GHz, that the absorption indeed is rising strongly at 7 T, indicating that the DA is approached. (The calculated peak position according to II could have been shifted also by adjusting the model parameters, but a parameter adjustment was not attempted. The absorption anomaly at 2310 GHz around ~ 6 T occurs therefore between a well defined hybrid and a well defined DA, which is also approximately true at 2523 GHz. At 2907 and 3106 GHz, the analysis is more complicated as the structure occurs here in the region where the smeared out hybrid is expected. Despite this complication, we believe the structure is caused by an effect which is in principle not included in our calculations: the features resulting from the calculations (a distinct hybrid or a gradual increase in absorption) are nicely present in the experiment, but the additional structure seems superimposed. The structure seems to scale with frequency (see Fig. 8c) and thus rules out possible complication with quantum oscillation effects. The line drawn in Fig. 8 does however not extrapolate through the origin. The strength of the structure appears to increase strongly with frequency (but note the scale change

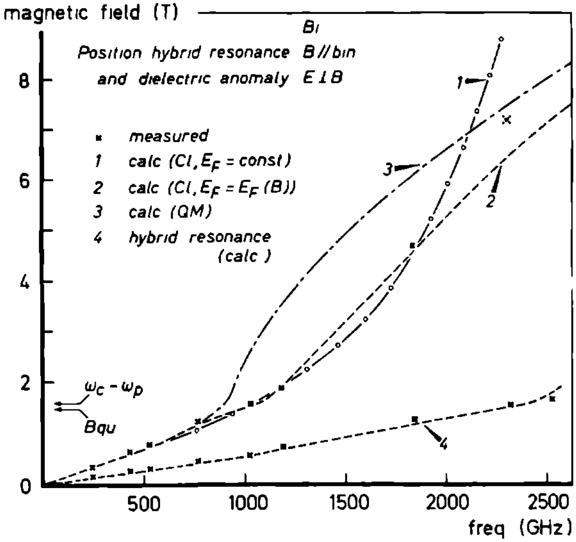


Fig. 9. Magnetic field positions of the hybrid resonance and associated dielectric anomaly occurring in the extra ordinary polarization as a function of frequency. Experimental results and calculations. The quantum limit field value B_{qu} is indicated by an arrow. Also indicated by an arrow is the field where the cyclotron-frequency of the light electrons, ω_c, equals the plasma-frequency, ω_p.

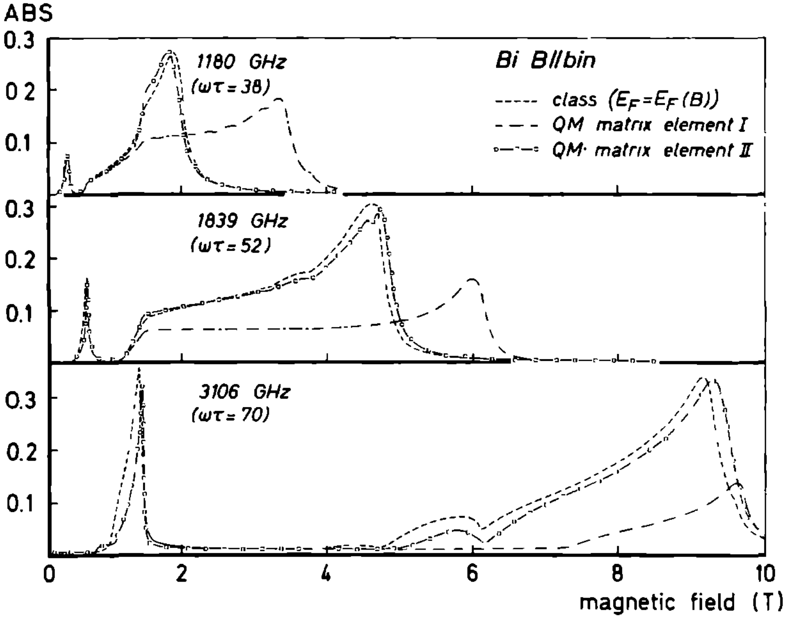


Fig. 10. Calculated absorption as a function of magnetic field in a frequency range below the plasmafrequency. The results obtained using two different multiplication factors in the matrix element of the two band model are compared with eachother and with the classical calculations."Matrix element I" employs eq. (18a) (Wolff) and "matrix element II" eq. (18c) (Hansen). Realistic ωτ-values are used (see text).

at 3106 GHz in Fig. 8a). The field region where it occurs is roughly around the third subharmonic of the hole cyclotron resonance ($\sim$ 6.3 T at 2523 GHz). This fact makes it tempting, but admittedly speculative, to associate the structure with nonlocal effects: the nonlocal wave dispersion has generally strong structure at the hole subharmonics[63,70], which would be reflected in the surface impedance.

Part of the results discussed so far are summarized in Fig. 9, where the field position of the electron-electron hybrid and the high field DA are plotted against frequency. For the experimental and calculated position, the hybrid resonance is taken to be the point of maximum curvature in curves such as those given in Figs. 4 and 8, while the DA is simply taken to be the maximum (although this maximum might be very broad and ill defined in case III). The field region enclosed between the lower and one of the upper curves in Fig. 9 defines a window, where (local) wave propagation can occur. At low fields and frequencies, lower and higher limits of the window scale linear with frequency and the experimental positions are in good agreement with the calculation. The DA measured for the corresponding orientation by Smith et al.[7] at 70 GHz, fits on the upper line. At fields $\gtrsim$ 1.5 T all curves calculated for the DA (as follow from I, II and III) start to deviate from the straight line behaviour. As in the classical formulas (I) no quantum limit is defined, this is evidently due to the fact that the cyclotron frequency approaches the plasma frequency. It is somewhat coincidental that the quantum limit is reached in the same field region where the cyclotron frequency becomes comparable to the plasma frequency (see the arrows in Fig. 9). The measured points agree satisfactorily well with the model II calculations. Accidentally and unfortunately, the measured point at 1839 GHz is near the crossing of curves I and II. The estimated point at 2310 GHz does not fit II, but in this region the results are rather sensitive to the model parameters, especially to the precise value of the plasma frequency, and thus n, $\varepsilon_1, \varepsilon_2$. This holds for model I as well. Calculated results following II, show a continuous, although slightly oscillatory, behaviour near B_{qu}, but the change for III is rather abrupt. The hybrid resonance fits a straight line rather well up to the field region $\gtrsim B_{qu}$, where it cannot be defined sharply. The difference in the calculation I, II and III would be hardly noticeable on the scale of Fig. 9.

As mentioned already, the strong discrepancy between the data and the quantum calculation III can be removed solely by choosing a different expression for the matrix element, or rather for the multiplication factor which is needed for the two band model (eq.(16)). This is explained in Fig. (10) where results are compared obtained by using either (18a) or (18c) as multiplication factors in (16). If the averaging factor of Hansen[51] (18c) is used, the quantumcalculations III reduce then almost to the semiclassical calculations II and are in view of the preceding results in excellent agreement with the data. Use of the (apparently incorrect[51]) expression (18b) would lead to curves intermediate between those obtained from (18a) and (18c). A further discussion of this issue will be postponed until the data at higher frequencies have been presented.

4 A iv. Cyclotron resonance.

We now consider the cyclotron resonance associated peak at low fields which is induced by the $\vec{E}//\vec{B}$ polarization mode . Attention will be devoted to the fundamental resonance only. A detailed analysis of the data reveals several new aspects of quantum cyclotron resonance. Especially, the pronounced influence of $k_B \neq 0$ contributions at high frequencies is very remarkable in view of the common sense knowledge that $k_B = 0$ contributions are predominant, as they profit from a maximum in the joint density of states between levels. $k_B = 0$ contributions are particularly significant in those cases where $k_B = 0$ transitions cannot occur due to the band scheme. The data will be analyzed using a purely local theory, including quantum effects however and possible nonlocal effects are discussed qualitatively. This should be compared to the semiclassical, nonlocal approach used in ref. 19, where local effects such as the DA were discussed qualitatively. The more rigorous treatment of Miclavc and Drew[50] is very complicated and necessarily approximate and was adapted to account for the subharmonics which result from a purely nonlocal effect indeed.

The resonance considered is a tilted orbit cyclotron resonance of the light electrons. Quantum effects enter in the line shape which is strongly affected by the nonparabolic band, while carrier redistribution effects are of secondary importance here as the resonance is due to a single group (two equivalent groups) of carriers.

The positions of the CR, such as can be obtained from the experimental data (e.g. Figs. 4-8) are plotted in Fig. 11a. In a direct absorption measurement, without field modulation, the resonance is defined as the onset of absorption which is well approximated by the point of largest curvature for not too small $\omega\tau$-values. The experimental points in Fig. 11a were determined in such a way, but neglecting some fine structure at the edge of the peak such as at e.g. 1839 and 3106 GHz. The straight line in Fig. 9 indicates the position which would be

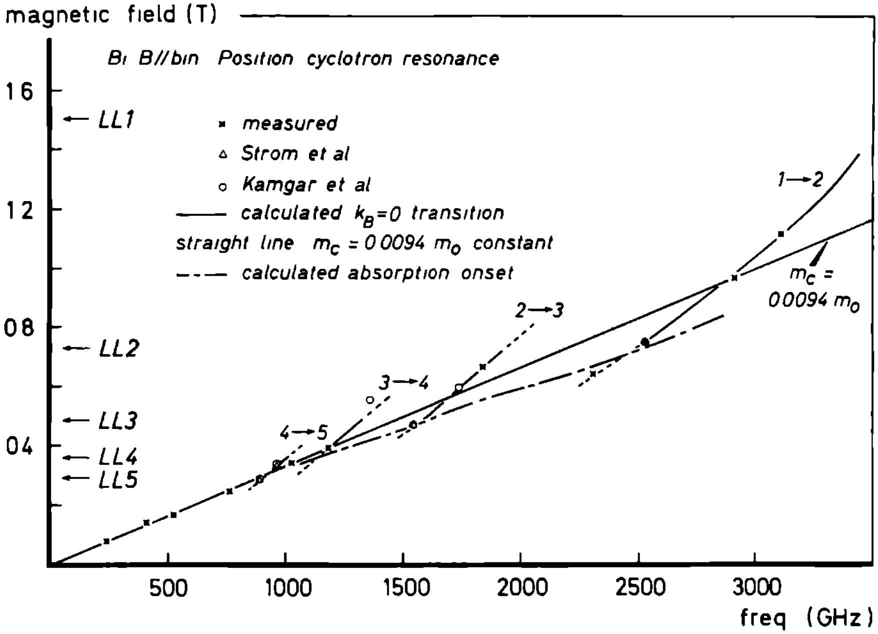


Fig. 11a. Magnetic field position of cyclotron resonance as a function of frequency. Straight line gives the position for a parabolic band with the low-frequency cyclotron mass of Bi. Calculated Landau level energy differences at k_B=0 are given by the curved lines, which are dashed in the regions where transitions are not compatible with the occupation factors. The short-long dashed line gives the absorption onset as calculated from the two band model, including $k_B \neq 0$ transitions. Field positions where the succesive Landau levels cross the Fermi level are indicated by arrows.

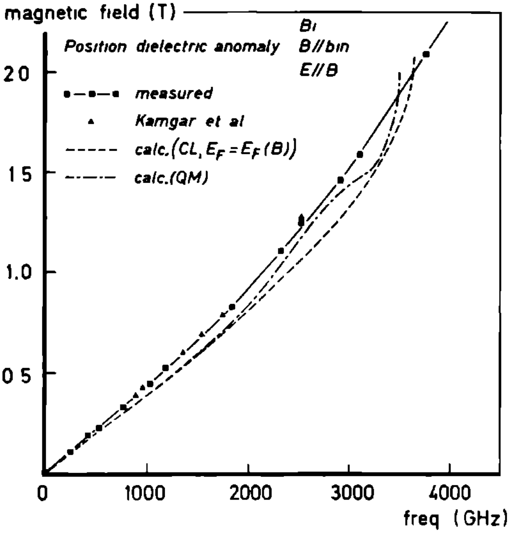


Fig. 11b. Magnetic field position of the dielectric anomaly resulting from the ordinary polarization (following the cyclotron resonance) as a function of frequency.

expected from a constant cyclotron mass while the curved lines correspond to the position of the k = 0 transition, calculated from (9).The last curves are drawn bold where k = 0 transitions are compatible with the occupation factors and dashed where these transitions cannot occur (see Fig. 3a). Field positions where Landau levels at k = 0 cross the Fermi level are indicated by arrows along the vertical axis. Data from Strom et al.[19] and Kamgar et al.[24] have been included also. It is seen that all experimental points indeed closely follow the calculated lines, supporting the view that $k \neq 0$ transitions are not too relevant. The nonparabolic band of Bi reflects itself thus strongly in FIR CR-experiments and the data are in excellent agreement with the two band model, as was already established before[19]. Note however that the 2310 GHz data point lies in the zone where k = 0 transitions are forbidden by the occupation factors.

In Fig. 11b the peak positions of the DA following the CR are plotted, together with the data of Kamgar et al.[24] (Actually, points taken from ref. 24 designate the point of steepest slope as obtained using derivative techniques; for high $\omega\tau$-values however, this point is sufficiently close to the peak position). The calculated positions according to the classical model II and the quantummechanical model III are also given. The difference between II and III is not yet very large, because the fields are essentially below the quantum limit. It was already noted by Kamgar et al.[24] that the calculated peak positions are consistently too low and this discrepancy extends down to the lowest frequencies and fields, so that it is not expected to be a quantum effect, in agreement with our calculations. As a possible cause, deviations of the Fermi surface from ellipsoidal was suggested[24]. Thus, this unexplained fact makes a detailed comparison of experimental and theoretical line shapes difficult. In contrast to the CR-position, the DA-position shows a very smooth scaling with frequency (although nonlinear at high frequencies). Some structure is present in the calculated curves near the quantum limit field B_{qu}, due to carrier density oscillations. Such structure seems also present in the experimental results, but the number of frequencies is too limited to show this more clearly.

Fig. 12 shows two typical CR-peaks in detail. Qualitatively, the most prominent features from the data, such as the small step at 1839 GHz and the large step at 2523 GHz , are well reproduced by the calculation (model III). Quantum lineshapes can be seen strongly to deviate from the classical ones, especially at 2523 GHz. At 1839 GHz, the main contribution to the resonance comes from the $k = 0, 2^- \rightarrow 3^-$ transition at $B \sim 0.66$ T. The lower level (2^-) is just below

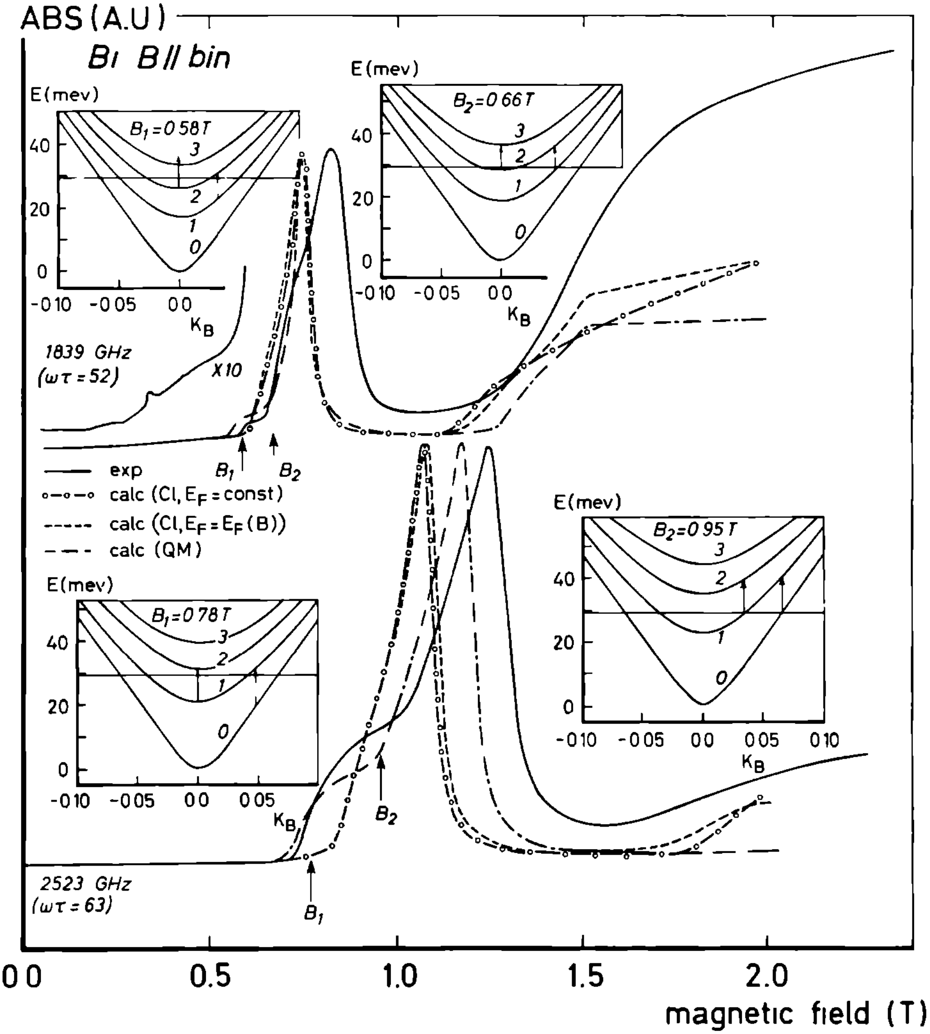


Fig. 12. Absorption lineshapes of cyclotron resonance and associated dielectric anomaly at two typical frequencies. Experimental results and calculations. Several insets illustrate the Landau level scheme applying to typical features in the (calculated) curves. At 1839 GHz (upper curve), B_1 points at the onset of absorption due to $k_B \neq 0$ transitions (left inset) and B_2 points to the central transitions (right inset). At 2523 GHz (lower curve), B_1 points at the onset of absorption, predominantly due to central transitions in this case (left inset) and B_2 points at the limiting resonance (right inset).

the Fermi level (see inset). At lower fields (∿ 0.58 T) $k \neq 0$ transitions between 1^- and 2^- are already important, and these cause the small step in front of the main absorption onset, as turns out from detailed analyses of the calculated line shapes. As level 2^- is not far from the Fermi level, the contribution $1^- \rightarrow 2^-$ is relatively large, but $2^- \rightarrow 3^-$ transitions occur in a limited range of k_B-values around $k_B = 0$ and extend over a small field range only. This results in a somewhat anomalously narrowed calculated peak. The observed behaviour is somewhat different. The small step, which should clearly be identified with the calculated $1^- \rightarrow 2^-$ transitions, is smaller than calculated, the $k = 0$ transition on the other hand, sets in much stronger. Note that the curvature of the measured peak near the main onset is of different sign than the calculated one. The different behaviour should certainly be attributed to nonlocal effects. Nonlocal effects are most important indeed near a resonance as follows from the local viewpoint ($q \rightarrow \pm \infty$ at resonance), as well as from explicit nonlocal calculations. The subharmonic resonance, clearly visible in absorption, using no field modulation, emphasizes the importance of nonlocal effects, while the nonlocal calculations of Miclavc et al.[50] for the subharmonics in a comparable frequency range, revealed explicitly that nonlocal type of wave propagation (cyclotron waves) strongly influences the subharmonic line shapes. Note also the gradual increase in absorption, visible in Fig. 12, between the second subharmonic and the onset of the fundamental at 1839 GHz. From Fig. 12 (1839 GHz), the effect of nonlocal wave propagation is apparently to emphasize the $k = 0$ transitions, leaving the $k \neq 0$ contributions relatively unaffected, or even to suppress them.

The situation is somewhat different at 2523 GHz. Now the CR is dominated by the $1^- \rightarrow 2^-$, $k = 0$ transition, the 2^- (upper) level being just above the Fermi level (see inset). This transition starts at ∿ 0.75 T, however it turns out that the $k \neq 0$, $0^- \rightarrow 1^-$ contributions (see inset) are nonnegligible and cause already some absorption at a field slightly lower than the onset of the $k=0$ transitions. When the field increases, resonant transitions take place at $k_B \neq 0$, both from $1^- \rightarrow 2^-$ as well as from $0^- \rightarrow 1^-$. This continues up to a field value indicated as B_2 in Fig. 12 (lower part), where both lower levels (1^- and 0^-) pass the Fermi level at the appropriate k_B-value, at nearly the same field value (0.95 T). (This situation should not be confused with the Landau level "resonances" mentioned before). The field value where $k_B \neq 0$ resonant transitions terminate, gives rise to a distinct kink in the calculated line shape (indicated by the arrow B_2 in Fig. 12). The same kink is also present in the experimental result, although it occurs at a somewhat higher field value . It is not clear, but interesting, whether the discrepancy in position of the kink is related to the discrepancy discussed earlier in the position of the Landau level resonances. In contrast to the small step at 1839 GHz, the large step at 2523 GHz is experimentally larger than calculated. Also the experimentally observed onset seems somewhat better defined than in the calculations (and starts at 0.75 T, corresponding to the k = 0 transitions). Nonlocal behaviour is again believed to be responsible for this; the apparent emphasis on the k = 0 transitions is consistent with the observations at 1839 GHz. A structure in the absorption around the field value B_2 (2523 GHz, Fig. 12) has not been observed or interpreted as such before. Its occurrence bears some resemblance with what in the language of microwave cyclotron resonance is called a limiting point resonance[3].

A glance at Fig. 8a, upper two curves, does not suggest that there would be a qualitative difference between the CR-peaks at 2310 or 2523 GHz. Yet, according to the band model used (Figs. 3a, 11a), the 2523 GHz CR can fully benefit from k = 0 transitions, while k = 0 transitions are not possible at 2310 GHz. Our local calculations do confirm this similar behaviour, which shows in a direct way the relative importance of $k \neq 0$ contributions as compared to k = 0 contributions Nevertheless, a careful comparison does show a qualitative difference; for this purpose the two line shapes are plotted in Fig. 13 with the field scale shifted such that the onset of absorption coincides in both cases. It is clearly seen then, that the rising slope at 2523 GHz is steeper than at 2310 GHz, a behaviour which is not present for the calculated results, also given in Fig. 13. The similarity of the high-field slopes of the experimental curves indicates that this is not likely to be a relaxation time effect (although a field dependent relaxation time cannot be ruled out in principle). Again, we believe that nonlocal effects emphasize the k = 0 transitions and account for the characteristic differences at the two frequencies.

The characteristic step seen at 2523 GHz becomes narrower as the frequency is increased (see Fig. 8a). The decreasing stepwidth is nicely reflected by the calculations also (Fig. 8a). Qualitatively, it is easily understood on the basis of level schemes such as given in the insets of Fig. 12 (lower part): the field separation between k = $0,1^- \rightarrow 2^-$ transitions and limiting resonances ($1^- \rightarrow 2^-$ or $0^- \rightarrow 1^-$, B_2 in lower part of Fig. 12) becomes smaller when the frequency (and so the resonance field) is raised. Because the upper level 2^- at 2523 GHz is just above the Fermi level, the step width is about maximal at 2523 GHz. Note also that at the three subsequent frequencies 2523, 2907 and 3106 GHz, the k = 0 transition is respectively below, almost coincident

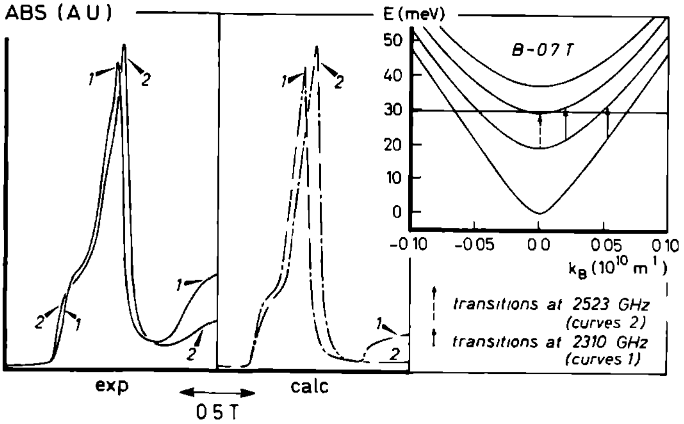


Fig. 13. Comparison of cyclotron resonance lineshapes at the two frequencies 2310 GHz and 2523 GHz, experimentally (left) and theoretically (right). For a detailed comparison, the absorption onsets at both frequencies have been made to coincide. The inset shows the relevant Landau level scheme. At 2523 GHz, central transitions are dominant (dashed arrow) and at 2310 GHz $k_B \neq 0$ transitions.

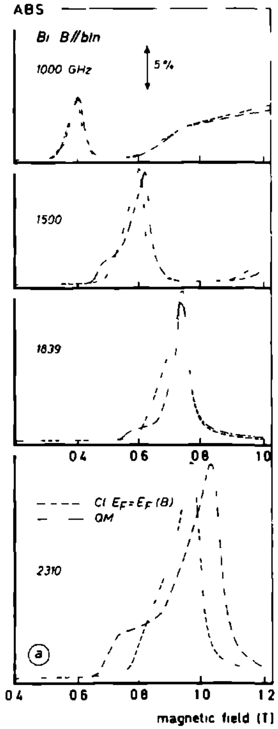


Fig. 14a. Calculated absorption lineshapes of cyclotron resonance and associated dielectric anomaly, at several typical frequencies in the intermediate frequency and field range.

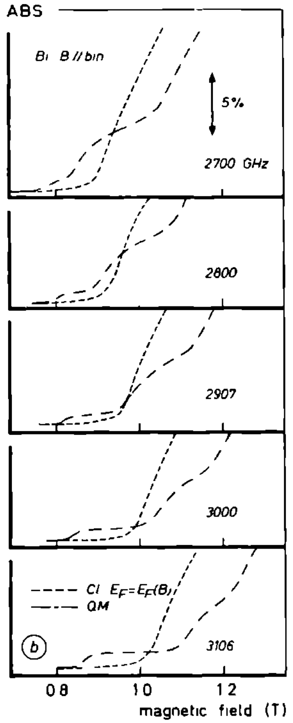


Fig. 14b. Calculated absorption structure associated with cyclotron resonance at high frequencies (but below ω_p).

with and above the classical CR position (see also Fig. 11a).

The theoretical behaviour of the CR line shapes is summarized in Fig. 14a. At the lowest frequency (1000 GHz) the line shape is practically classical but a foot structure starts to develop. At 1500 GHz a pronounced step is present resulting from $k \neq 0$ transitions $2^- \rightarrow 3^-$, $1^- \rightarrow 2^-$, the $k = 0$ transitions being forbidden (see Fig. 11a); the second (upwards) inflection point corresponds to a limiting resonance . The step at 1839 GHz results again from $k \neq 0$ transitions ($1^- \rightarrow 2^-$) but now the second (upwards) inflection point indicates the onset of $k = 0$ ($2^- \rightarrow 3^-$) transitions. Because the lower level (2^-) is close to the Fermi level, (see Fig. 12 insets), the limiting resonance is close to the $k = 0$ resonance as discussed, and therefore the peak is anomalously narrow. Such a sequence repeats itself at higher frequencies, with lower index Landau levels. The step features are best resolved at the highest frequencies, where the lowest index Landau levels are involved. This is shown in Fig. 14b, where the development of a two step structure is depicted. The first step extends from the onset of the $k \neq 0$ transition ($0^- \rightarrow 1^-$) to the onset of the $k=0$ ($1^- \rightarrow 2^-$) transition. The second step starts at $k = 0$ ($1^- \rightarrow 2^-$) transitions and ends at a limiting resonance. Although the $k = 0$ transition may be at a field value above or below the classical position (see Fig. 11a), the absorption onset (including $k \neq 0$) always starts below the classical value. Its calculated position is also given in Fig. 11a. Note that the 2310 GHz point accidentally lies close to the (forbidden) $1^- \rightarrow 2^-$ $k = 0$ transition.

In the experimental curves, the $k \neq 0$ step is not seen at 2907 GHz, but clearly present at 3106 GHz. The 3106 GHz data are shown in more detail in Fig. 15, together with some clarifying insets. Also at this frequency, it is immediately seen in Fig. 15 that the experimental $k \neq 0$ step is much smaller than expected from the calculations, while the $k = 0$ transition results in a much steeper rising slope than calculated. This behaviour is perfectly consistent with the observations at the lower frequencies. A direct evidence for the importance of nonlocal effects even at these high frequencies is seen in the occurrence of the subharmonic at 2907 GHz (Fig. 8a). It should be stressed that the apparent prominence of the $k = 0$ transition cannot be explained solely from the maximum in the joint density of states at $k_B = 0$, as this is properly incorporated in the theory.

In most of the experimental curves (see Fig. 8a, Fig. 15) an inflection point is present close to the top. This structure is not present in any of the calculated curves, also not when splitting is included. It is therefore not expected to be a bandstructure effect (at least within the band model used).

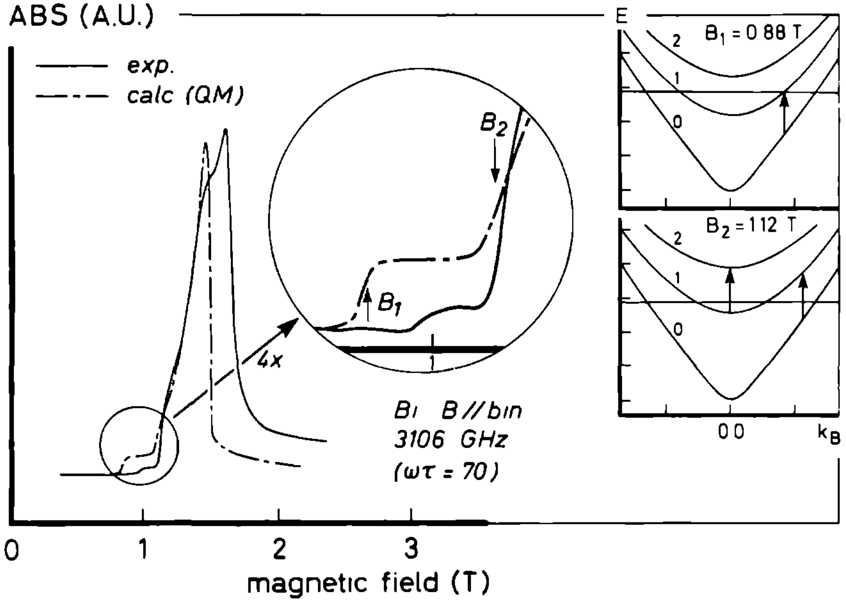


Fig. 15. Absorption lineshape of cyclotron resonance and associated dielectric anomaly at 3106 GHz. Enlarged view shows a detail of the absorption onset with a step structure due to $k \neq 0$ transitions. Experimental results and calculations. The insets show the Landau level schemes applying to the field values indicated by B_1 and B_2 in the calculated curves.

The structure is neither caused by a quantum oscillation effect, as it moves with frequency and is always located close to the top of the peak. Moreover, it occurs well above the limiting resonance, so that resonant transitions would not contribute to the line shape in this part of the peak. The present data however are insufficient to justify a serious attempt to assign the structure. An investigation of the orientational dependence would be highly desired to rule out an orientation effect. In this respect, the discrepancy in positions, which does exist also for those structures ($k \neq 0$ resonances, limiting resonances), which could be unambiguously assigned, should be remarked. These structures, which depend upon the k_B-dependence of the energy dispersion are expected to depend stronger on orientation than the central orbit ($k = 0$) resonances. Moreover, the known deviations from ellipsoidity of the Fermi surface[3], might also influence such k_B-dependent structures.

In view of the strong influence, details of the bandstructure do have on the line shapes, as shown in the preceding, great care should be taken in relating anomalies in the line shape with peculiar points of the dispersion relation of nonlocal waves, which was attempted in ref. 19.

4 A v. Behaviour for $\omega \gtrsim \omega_p$

We will now discuss how the magnetoplasma behaviour as discussed up to now evolves when the applied frequency approaches, and eventually exceeds, the plasma frequency. In order to do so, it is most instructive to sketch the calculated behaviour first. As now both polarization modes obtain strong structure in the same field region, it is necessary for a qualitative understanding to present the calculated results for both polarizations separately. In Fig. 16a the behaviour is given for frequencies just below the plasma frequency (∿ 4700 GHz). We will consider the behaviour only in the field range corresponding to the experimental. It should be remarked however, that at these frequencies, structure in the absorption may be present up to very high fields ($\gtrsim$ 30 T) due to heavy mass carriers (notably the holes).

Characteristic feature for the CR-associated absorption (E//B) is that the DA, following CR gradually disappears and the medium becomes transparent at frequencies above CR. In the transition region, the absorption shows extremely strong hole quantum oscillations, which is expected for an effect depending strongly on the plasmafrequency (and thus carrier density). The classical (I), modified classical (II) and quantum model (III) behave qualitatively similar. The quantum model has considerable structure at the leading edge. The

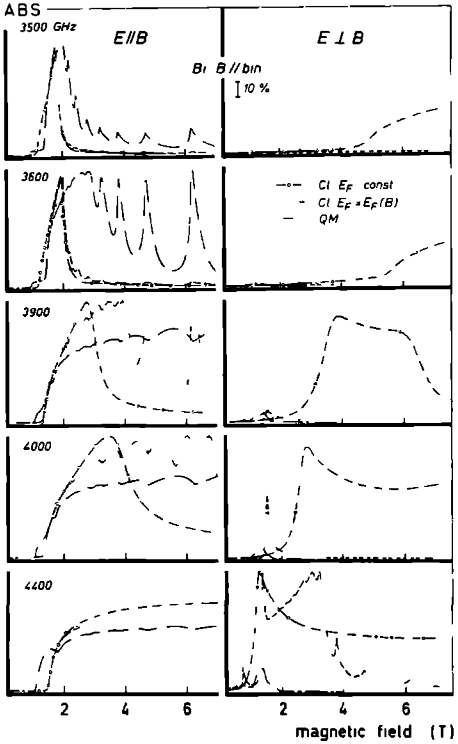


Fig. 16a. Calculated absorption as a function of magnetic field for frequencies just below the plasma-frequency. Left column: ordinary polarization; right column: extra ordinary polarization. Realistic ωτ-values are used (see text).

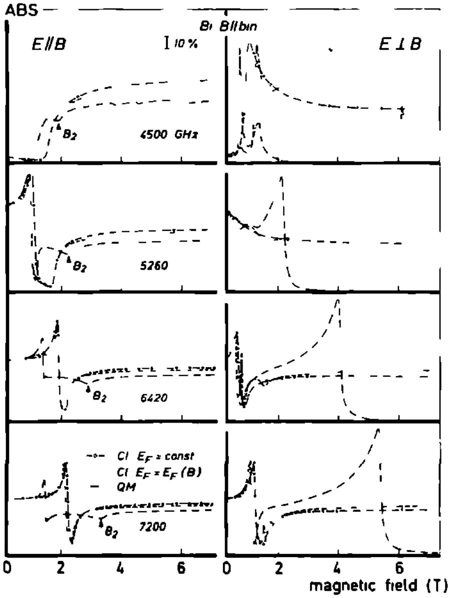


Fig. 16b. As Fig. 16a, but now for frequencies near and above the plasmafrequency.

typical frequencies at which the region above the DA becomes transparent, differ by at least 10% for the models II and III.

For the perpendicular polarization, at the lowest frequencies of Fig.16a, the electron-electron hybrid is only present in the classical model I. As discussed before, the increase in carrier density at high fields has moved it outside the field range displayed, in the other two models. At higher frequencies, it is shown how a so called field dependent plasma edge[2] starts to develop. It starts at a field near CR and is cut off at a higher field. The onset remains fixed near CR, but the cut off rapidly moves to higher fields. There are large quantitative differences between all three models. The plasma edge in model III is only just seen at 4400 GHz.

Fig. 16b shows the development when the plasma frequency gets exceeded. The CR-absorption onset for E//B remains relatively unaffected, but a zero field absorption develops above the plasma frequency. The zero field absorption is cut off again well below the CR field position. Such a cut off, which reflects a zero in the effective dielectric constant, is also known as dielectric anomaly in the literature. A cyclotron resonance in the Voigt configuration is always accompanied by a DA. Because at resonance the DA goes to ± ∞, it necessarily passes through zero. Depending on whether the frequency is above or below the plasma frequency (ε positive or negative at 0 field), the DA falls just below, or just above the CR.

The position of the DA and associated CR in model III differs considerably from model I and II and is easily understood in terms of the nonparabolic band. The anomalous feature marked B_2 in Fig. 16, corresponds to a limiting resonance. For E ⊥ B, a similar DA and resonance develop. The positions however are different for E//B and E ⊥ B. The absorption cut-off following the resonance is a distinct feature of the quantum model, and has moved far outside the displayed field range for the other two models. It will be shown that the markedly different behaviour can be accounted for by the matrix element used. Close to the plasma frequency, quantum oscillations due to the light electrons can be seen at low fields, most clearly for the perpendicular polarization. The qualitative behaviour as sketched in Fig. 16 can be understood from the classical model I and is extensively discussed in standard textbooks (e.g. ref. 2).

The experimental results at two frequencies just below the plasma frequency are given in Fig. 17. For comparison, some calculated curves are given also. Comparison with the quantum calculations should be done with some care, because there is a great arbitrariness in scaling. As explained in the inset, a k = 0 resonance is not possible at 3747 nor at 4252 GHz. Instead, absorp-

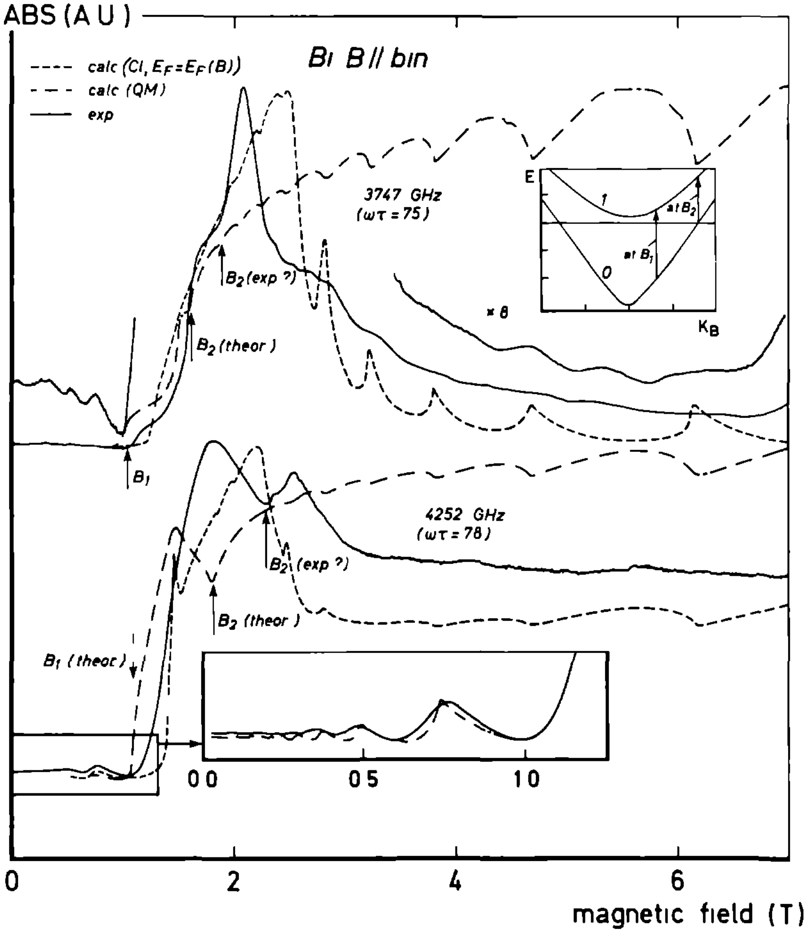


Fig. 17. Absorption as a function of magnetic field for two frequencies just below the plasmafrequency. Experimental results and calculations. The inset shows the relevant Landau level scheme, explaining special field values designated by B_1 and B_2. The magnifications show quantum oscillations of the carrier density with periods corresponding to the light electrons (low field) and holes (high field).

tion starts to rise at some field value designated by B_1 where $k \neq 0$, $0^- \rightarrow 1^-$ transitions can take place. The calculated curve shows again a discontinuity at a field value B_2 which corresponds to a limiting resonance. The nonmonotic behaviour between the two values B_1 and B_2 is related to the Landau level resonance 1^- of the light electrons. The leading slope resembles somewhat that at lower frequencies (e.g. 3106 GHz. Fig. 8a and 15), the structure between B_1 and B_2 however is of different origin. Qualitatively, the leading edge of the experimental result is in good agreement with the calculated result (III); the two-step structure is strongly present. Agreement in field positions is evidently poor, especially as regards the limiting resonance. The 1^- Landau level resonance at B_{qu} is experimentally too high, but consistent with our observations at lower frequencies. At the high field side, the experimental results are in very reasonable agreement with model II, but clearly not with model III. Especially the hole quantum oscillations are smaller than calculated and smeared out, but clearly present. The experimental period spacing,which seems to disagree with the calculation, can be explained by the spin splitting of the hole levels[44]. Note the onset of absorption at the highest fields at 3747 GHz, not present in the calculations, which might be related to the unidentified high field structure in Fig. 8b, c.

The absorption spectrum at 4252 GHz is very similar. As this frequency is still closer to the plasma frequency, the DA following CR is no longer present and the parallel polarization is expected to show a step only (see Fig. 16a). The perpendicular polarization now has also a resonance in model II and a cut-off. As a result of averaging over both polarizations, the calculated absorption in model II shows a decrease in absorption in Fig. 17, corresponding to the cut-off in the perpendicular mode. Such a cut-off is clearly also displayed by the experiment. Following model III, the perpendicular mode does not yet contribute to the absorption at 4252 GHz (see Fig. 16a), in clear contrast with the experiment. On comparison with the calculation, the dip between the two peaks should be attributed again to a limiting resonance (marked B_2 for the experimental and calculated curve; see also inset). The measured and calculated positions of the limiting resonance differ greatly; the experimental value might in this case however be incorrect as it might be apparently displaced due to the combined effects of both polarizations. The difference in onset of the absorption is related to the position of Landau level 1^- with respect to the Fermi level in both cases. The small step at 3747 disappears gradually, as can be seen in Fig. 16a. The cut-off field value at 4252 GHz is very sensitive to the plasma frequency and could have been fitted by adjusting model parameters (e.g. n or ε_i); the discrepancy is therefore not significant. The calculated positions of CR (onset peak) and limiting resonance vary by at most 5% when the model parameters change within reasonable limits (i.e. introducing the correct spinsplitting and variation of the FL, but keeping masses at the FL fixed). The relative difference between the assigned values B_2 at the two frequencies is about twice as large at 4252 GHz (∿ 20%) than at 3747 GHz (∿ 10%), which indicates also that the apparent position at 4252 GHz might be distorted by the other polarization mode. Although the recordings in Fig. 17 were accidentally taken with a time delay of 2 weeks in between, the sample was nominally untouched and thus the orientation unchanged. A misorientation of the sample is presumably at least partly responsible for the discrepancy in position, because the positions are determined by $k \neq 0$ transitions and depend on m_z also, which depends rather strongly on orientation in this direction.

The spectrum at 4252 GHz shows rather strong quantum oscillations of the light electrons at low fields, shown on an enlarged scale in the inset. Similar oscillations are seen at 3747 GHz, although of much smaller amplitude. Such type of oscillations were first observed by Dresselhaus et al.[72] in Sb. They are usually referred to as "optical Shubnikov de Haas effect". Clearly, their origin is similar to the origin of the other quantum oscillations (due to holes), observed at other frequencies and other field regions. As now the frequency is close to the plasmafrequency, the zero field absorption is strongly dependent on ω_p and thus on carrier density and small variations of carrier density are strongly reflected in the absorption. The strong frequency-dependence of the amplitudes can thus be explained. The order of magnitude of the amplitude is well in agreement with the calculated magnitude (if reference is made to the main peak magnitude) and they arise strongest for the $E \perp B$ polarization. For the present experiment, there is no direct need to invoke other mechanisms than simple carrier density oscillations, to explain the presence and magnitude of the oscillation. On the other hand, the low field oscillations appear much stronger than other quantum oscillations, e.g. the hole oscillations in the high-field region at 3747 GHz (Fig. 17), which are much smaller than calculated. It should be kept in mind however that the calculations were done neglecting the level broadening, which suppresses and smears out the quantum oscillations. The level broadening will be of order $1/\omega_c\tau$ and assuming equal τ's for electrons and holes, $\omega_c\tau$ is much smaller for holes than for (light) electrons, even if the holes are in the high field range of Fig. 17.

In a more detailed study of the optical ShdH-effect in Sb, Missel et al.[34]

found that the amplitudes experimentally were orders of magnitude larger than expected from simple density oscillation models and that these also were only weakly frequency-dependent. The discrepancy from the simple theory was discussed in terms of a magnetic field dependence of the relaxation time. Introduction of a field (or n and k_z) dependent relaxation time in the high-frequency conductivity is nontrivial, but some explicit calculations were given by Mase[73], who actually also gave some typical examples as applied to Bi. He found that quantum oscillations may become very large at frequencies very near the plasmafrequency and are much larger than in the Shubnikov de Haas experiments. This theory did neither explain all the experimental observations in Sb[34] and it was suggested that (non resonant) interband contributions are essential for their observations. A ShdH-effect in the optical properties, frequently met in degenerate semiconductors (PbS[74], InSb and PbTe[75], Te[76], Cd_3As_2[77]), is attributed to nonresonant (i.e. non absorbing) interband contributions to the optical conductivity. Such effects become important when the frequency approaches an interband edge frequency. Interband contributions clearly are of no importance for the present case of Bi.

A Shubnikov de Haas-like effect in the optical properties similar to the one discussed for Bi is also possible for a multivalleyed band with fixed density. Such a case was met in n-PbTe[78]. In a magnetic field, the density of states of nonequivalent carrier pockets becomes different and thus the relative population changes. The ShdH-oscillations, which resulted for n-PbTe were termed "charge transfer resonance"[78]. Such a mechanism is in principle also active for Bi; in a magnetic field the population of the B, C-pockets becomes different from the population of the A-pocket. The relative population difference is maximal at the light electron LL-resonances, as a direct calculation easily shows (the population of B, C-being a minimum, the one of the A-electrons being a maximum). The light electrons effectively contribute stronger to the optical conductivity than the heavy electrons and thus a ShdH-effect would be expected. For the case of Bi however, the total density changes closely follow the density changes of the B, C-pockets and are almost as large. The so called charge transfer resonance is therefore of secondary importance to the optical ShdH-effect as compared to the total density oscillations, which are possible for semimetals only.

At very low frequencies, quantum oscillations are also frequently observed in metals, and have a similar origin as the dc ShdH-effect in metals (change in scattering time as the density of states at the Fermi level is changed by the magnetic field). They are observed in the quasistatic regime ($\omega\tau < 1$)[79],

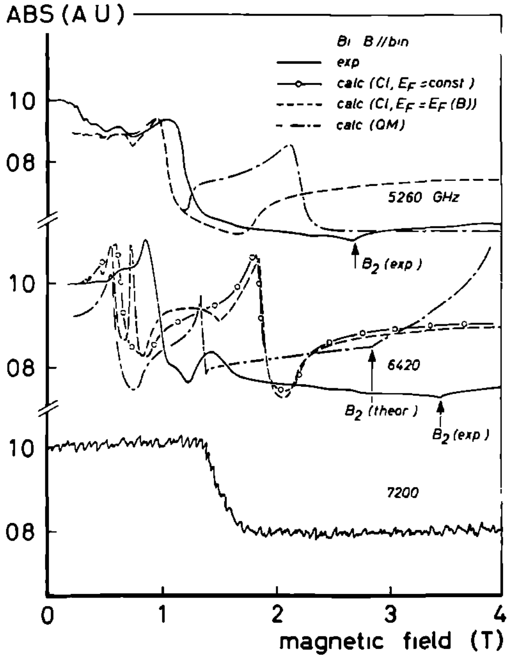


Fig. 18. Absorption as a function of magnetic field at three frequencies above the plasmafrequency. Experimental results and calculations. Structure designated by B_2 corresponds to limiting resonances (see inset in Fig. 17).

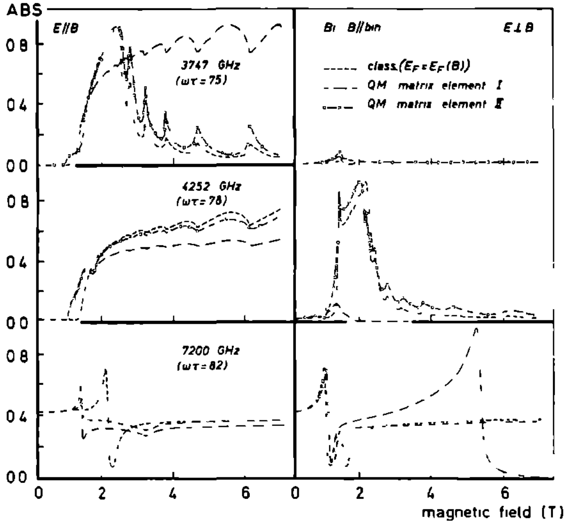


Fig. 19. Calculated absorption as a function of magnetic field in a frequency range near and above the plasmafrequency. The results obtained using two different multiplication factors in the matrix element of the two band model are compared with eachother and with the classical calculations. "Matrix element I" employs eq. (18a) (Wolff) and "matrix element II" eq. (18c) (Hansen). Realistic ωτ-values are used (see text). (Compare with Fig. 10).

predicted[80] and observed[81], in the high-frequency ($\omega\tau > 1$, but $\omega \lll \omega_p$, $\omega_{\text{interband}}$), local ($R_c < \delta$) regime and observed also in the nonlocal regime ($\omega\tau>1$, $\omega \ll \omega_p$, $\omega_{\text{interband}}$, $\omega < \omega_c$, $R_c > \delta$)[37].

The quantum oscillations observed by Giura et al. are obviously of entirely different origin[57].

Although the occurrence of the optical ShdH-effect is a very common phenomenon, it is interesting and important to identify its cause in view of the many distinctly different physical mechanisms which all lead to the same experimental phenomenon, and sometimes cause confusion in the literature. The oscillatory behaviour observed in the present work in Bi, appears to be satisfactorily explained by a simple mechanism proposed for the first time by Dresselhaus et al.[72]. As all quantum oscillation effects observed in this work (not only the low field oscillations) are associated with other absorption phenomena, depending strongly on carrier density or FL positions, they bear in this respect a similarity with the type of quantum oscillation observed in the line shape of the CR of two-dimensional space charge layers[82].

We will now discuss the results obtained in the highest frequency range obtainable, which are displayed in Fig. 18. The frequency is now above the plasmafrequency, and the results show the characteristic plasma resonances, whose theoretical behaviour is given in Fig. 16b. At 5260 GHz, a drop in reflection occurs around 1T, which should on comparison with Fig. 16b, be attributed to the DA in the parallel mode. The subsequent CR-onset is apparently not resolved, but is very close to the DA according to III. For the CR, a similar situation applies as sketched in the inset of Fig. 17, i.e. $k = 0$ transitions are not possible, but CR starts at some value $k \neq 0$ (level 1 may be below the FL however, for fields $\lesssim 1.5$ T). The experiment shows no indication of the cut off around 2 T, expected in model III for the perpendicular polarization. A kink in the absorption around 2.7 T and marked with an arrow B_2, is however clearly visible. It corresponds to the limiting resonance ($E//B$) (see Fig. 16b), but is in the theoretical curve III in Fig. 18 disguised by the $E \perp B$ cut off.

At 6420 GHz, the absorption decrease seems to have moved to lower fields, but should be identified with the DA for $E \perp B$, which has appeared at this frequency (compare also Fig. 16b). A small peak is visible around 1.5 T, whose position corresponds well with the DA in the parallel mode in model III. Structure between the DA's is experimentally hardly resolved. In contrast to the 5260 GHz case, the position of the $E//B$ DA depends on whether a classical or a quantum model is used. The experimental data seem to favor the QM-model. Also at this frequency, the E ⊥ B cut off in model III is not observed experimentally (neither in the field range 4-7 T, where the absorption was flat and therefore not is displayed in Fig. 18). The limiting resonance also is clearly present at this frequency. Its position deviates considerably from the calculated position, but the deviation is consistent with that observed at lower frequencies.

The data at 7200 GHz are similar to those at the preceding two frequencies, but are of less quality (due to a less stable laser at this very short wavelength) and fine structure is not resolved.

Measurements of the type given in Fig. 18 were reported already by Hebel and Wolff[83] in the frequency range between 5500 and 6500 GHz. The resonances were interpreted as $k \neq 0$ cyclotron resonances within the two band model; the limiting resonance was not reported in that work.

The plasma effects observed in the high-frequency range (Fig. 17 + 18) show considerable deviations from the quantum model III. This concerns the high-field absorption for $\omega \lesssim \omega_p$ (Fig. 17) and the E ⊥ B cut off for $\omega > \omega_p$ which is not present in the experimental result. The discrepancy can be fully explained by the expression for the velocity matrix elements as shown in Fig. 19, where the calculated results are compared, obtained by using either the matrix element from Wolff[49] (eq. 18a) or from Hansen[51] (eq. 18c). With the latter matrix element, the quantum model largely reduces to the quasi-classical model II, however retaining the peculiar quantum structure associated with the cyclotron resonances. Thus all of the qualitative discrepancy between the data and the quantum calculation at high frequencies (Fig. 17 and 18) is removed, if expression (18c) is used in the matrix element. This conclusion is perfectly consistent with the conclusion reached in discussing the low-frequency data (see discussion around Fig. 10). It should be remarked that the apparent failure of expression (18a) is not expected to be due to the neglect of a second part of the matrix element given by Wolff[49], because the same simplifications are implicitly made in arriving at (18c) (or 18b).

4 B. Bi, B // bisectrix axis.

The case with the magnetic field along a bisectrix axis in the trigonal plane will not be treated extensively, because the qualitative features are similar to those with B along a binary axis. Data have been taken at low fields only. The field is almost (apart from the 6.5° tilt out of the trigonal plane) along the longest axis of one electron ellipsoid (A-electrons), which yield a very low cyclotron mass. The other two pockets (B,C-electrons) are inclined with respect to the field over roughly 60^o (apart from the tilt). The B,C-electrons ("heavy") also yield a light cyclotron mass which is only a factor of two higher than the A-electron mass. There is less symmetry in this direction as compared to the binary direction, which makes that all of the tensor components differ from zero. As a consequence, calculation of the absorption is highly involved[7] and the results more difficult to interpret physically. A true distinction between ordinary and extra ordinary modes cannot be made because mode mixing occurs in the sample due to the complicated conductivity tensor.

The calculations were performed for the external radiation polarization parallel and perpendicular to B following Smith et al[7]. As now both electron masses of the electrons are relatively light, both sets of carriers are quantized at lower fields. The A-electrons reach the quantum limit at ∿ 1.2 T and the B, C-electrons at ∿ 2.5 T. Besides the contribution of the A-electrons, we have for the bisectrix direction also the contribution of the (heavy) B,C-electrons calculated quantummechanically as well, although all experimental data were taken below 2.5 T, the quantum limit for the B, C-electrons.

Fig. 20 displays the experimental results obtained in the low-frequency region and the comparison with the calculations. The spectra are very similar to those obtained for the binary direction (compare with Fig. 4); even the peak positions are nearly equal, only the high-field hybrid, which can be seen at low frequencies only, occurs at a lower field value now. Despite this great similarity, the interpretation should be entirely different according to the theory. The strong, high-field peak is the DA associated with the tilted cyclotron resonance of the B, C-electrons and is expected for the E//B polarization. The low field peak is the DA expected between the first hybrid (onset of the small peak) and the second hybrid (seen at the three lowest frequencies of Fig. 20) and occurs for E ⊥ B. A tilted orbit CR and DA would be expected from the A-electrons just in front of the small peak of Fig. 20, but is very small because of the very small tilt (6.5^o) and is not seen in the calculated curves of Fig. 20. According to the general discussion of Smith et al.[7], for E ⊥ B, CR of the B, C-electrons might be expected also because the two orbits (B and C) are nonequivalent, but they are not present in (polarization resolved) calculations (see also Fig. 11 of ref. 7); actually they seem to be present however in the polarization resolved microwave data of Smith et al.[7], see quoted Figure. The mode mixing mechanism prevents a qualitative discussion

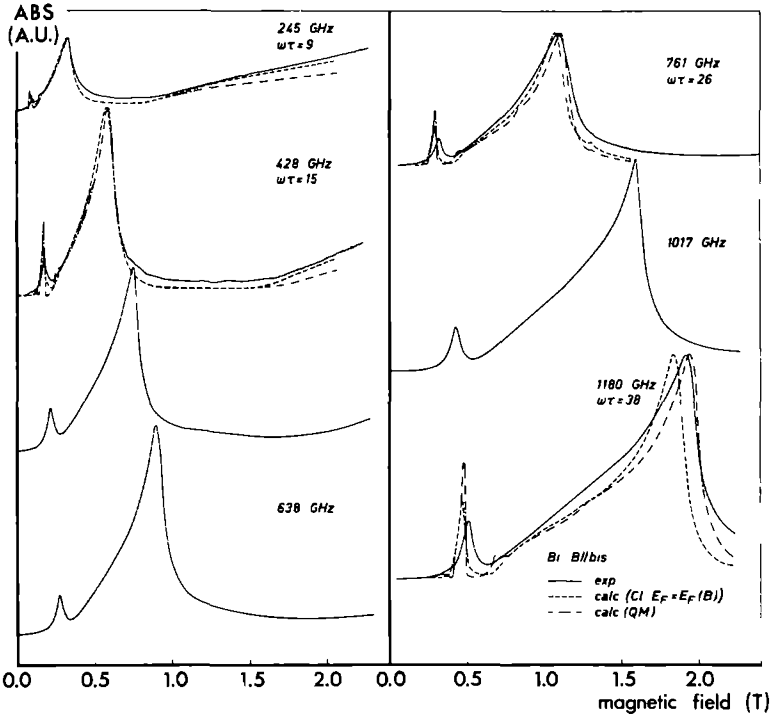


Fig. 20. Absorption as a function of magnetic field B for B along a bisectrix axis in the low frequency region. Calculated results are given for some frequencies.

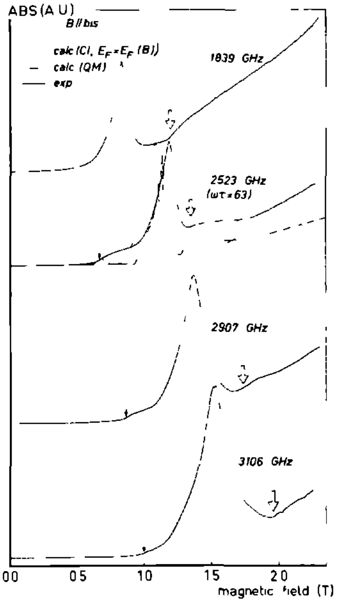


Fig. 21. Absorption as a function of magnetic field B for B along a bisectrix axis. The main peak is a dielectric anomaly and the arrows point at structure near the cyclotron resonances of the light and heavy electrons.

in terms of parallel and perpendicular polarized modes in the sample.

For the calculated curves, the same relaxation times were used as for the binary case . It can be seen from Fig. 20 that they apply very well also for this orientation. All the curves in Fig. 20 are in very good agreement with the simple classical model (II). Structure due to Landau level resonances, especially at the quantum limit of the A-electrons (1.3 T), is not observed in the data, in contrast with the results for B//bin. Such a structure however is also hardly present in the calculations. The difference between the classical model II and the quantum model III is very small, even in the quantum limit of the A-electrons (see curves at 1017 and 1180 GHz), again in sharp contrast with the data and calculations for the binary orientation. The major reason for this classical behaviour evidently is that the A-electrons have a less crucial contribution to the total conductivity as compared to the light electrons for the binary orientation. The BC-electrons now have only a ∿ 2 times higher mass, but also a 2 times higher density than the A-electrons. It was verified that the quantum calculations are practically invariant to whether expression (18a) or (18c) is used in the matrix elements. In the experimental results at the highest frequencies, a step-like structure can be seen just in front of the small peak. This can be attributed to the CR of the A-electrons. The quantum calculations do show absorption at the A-electron CR, but it is hardly present in the classical calculations.

The behaviour of the small peak and the region around it at higher fields and frequencies is displayed in Fig. 21, together with the calculated results at 2523 GHz. Also in this field and frequency region, the hybrid resonance and associated DA peak behave very classical. To avoid confusion, it should be emphasized again that the peak observed here is of entirely different origin than the peak observed for the binary direction , in the same field and frequency region (compare with Fig. 12). At these higher frequencies, the cyclotron resonances are better resolved. The onset of the A-electron CR as well as the B, C-electron cyclotron resonance is clearly seen and marked by a single and double arrow respectively in Fig. 21. The A-electron CR appears very much stronger than expected from the calculation. This might be due to nonlocal effects (and to some extent certainly is, in view of the findings for the binary direction), but a misorientation is expected to play an important role also. Because it is a tilted orbit resonance of an orbit which has only a slight tilt for ideal alignment, a misorientation will have an important influence on the strength, rather than on the position of the resonance. The plateau in the absorption which follows after the onset of the B, C – CR at

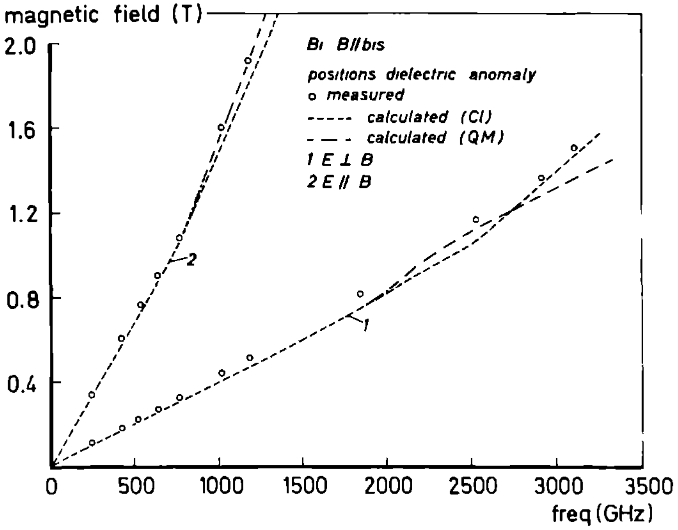


Fig. 22. Magnetic field positions of two dielectric anomalies as a function of frequency for the magnetic field along a bisectrix axis. The calculated results for E ⊥ B and E // B were obtained from the expressions of ref. 6.

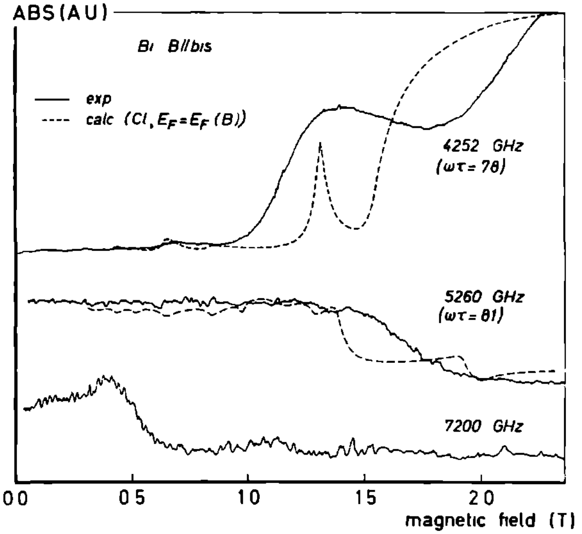


Fig. 23. Absorption as a function of magnetic field B for B along a bisectrix axis for frequencies near and above the plasmafrequency.

2523 GHz, qualitatively is nicely reflected by the quantum calculations. The end of the plateau is evidently determined by a limiting resonance, as discussed extensively for the binary orientation.

The peak positions of the two DA's as observed for the bisectrix direction are plotted against frequency in Fig. 22. The low-field peak deviates less than 10% from the calculated results and the high-field peak fits the calculations very well. Little difference exists between the classically and quantummechanically calculated position (except in the highest frequency range of the low-field DA). Deviations from straight line behaviour set in for fields around 1T and result from the magnitude of the cyclotron frequency which becomes comparable to the plasma frequency (see discussions of the corresponding Figures 9 and 11 for the binary case).

To conclude this section on the bisectrix direction, some data at frequencies very near and above the plasma frequency are given in Fig. 23. The qualitative features are highly similar to those discussed for the binary orientation (compare with the corresponding Figures 17 and 18). The complicated procedure to calculate the absorption from the conductivities was very sensitive to values obtained for them in the quantummechanical model, where the summation (13) contains resonant terms over much of the range displayed in Fig. 23, arising from $k \neq 0$ transitions of both A-and B, C-electrons. The numerical method of evaluating (13), used successfully in all other situations, turned out to give somewhat unreliable results in the cases of Fig. 23, and the quantum calculations are therefore omitted from the figure. The large rise in absorption around 1.1 T at 4252 GHz results from $k \neq 0$ cyclotron resonance of the A-electrons, while the rise near 2 T indicates the onset of the hybrid . In the region $B \gtrsim 2$ T, $k \neq 0$ cyclotron resonance also takes place, but cannot be identified from the data. Some quantum oscillations due to electrons (light and heavy) can be seen around 0.7 T. The peak displayed by the classical calculations at $\sim$ 1.3 T is a quantum oscillations at the quantum limit B_{qu} of the A-electrons. In the experimental curve, it is obscured by the cyclotron resonances. The two frequencies 5260 and 7200 GHz, show two high-frequency ($\omega > \omega_p$) dielectric anomalies, apparently of different origin in both cases. These were not analyzed further, but a similar discussion applies as given in connection with Fig. 18.

4 C. Sb.

The element Sb is found just above Bi in the same column of the periodic table, and is very similar to Bi in many respects. It has the same crystal structure and thus the same Brillouin zone (BZ), and is semimetallic also. Yet, there are several characteristic differences which make it exceptionally useful for investigation next to an investigation of effects occurring in Bi. In this respect it should be regarded as providing a means for changing the material parameters of the sample in a drastic way.

The geometry of the electron-FS is very similar to that of Bi, but the hole-FS is rather different. As compared to Bi, the hole pocket is slightly displaced from the symmetry point T in the BZ and because of this less symmetrical situation in the BZ, there are six hole pockets whose shape and orientation also is less symmetrical than the Bi hole pocket. The shape of both electron and hole pockets is known to deviate rather strongly from ellipsoidal[84], the deviation being somewhat stronger for the holes than for the electrons. The pockets are nevertheless usually approximated by ellipsoids, and this simplification will also be made in the present work. Carrier density is about two orders of magnitude higher than in Bi ($5.5 \cdot 10^{19}$ cm^{-3} [84]) and consequently, the Fermi energy also is much higher (100 meV), so that Sb is more metallic in character than Bi.

Because of the high Fermi energy and relatively high effective masses, quantum effects in the conductivity can be completely ignored in the field range considered and so far the experiments on Sb are not related to the main issue discussed in the present work . On the other hand, nonlocal effects, which were shown to play an important, although somewhat poorly defined role in the results on Bi, even at high fields, are expected to be stronger in the case of Sb. The plasma frequency of Sb ($\sim$ 30 THz) is almost an order of magnitude higher than of Bi and thus the (high-frequency, local) skin depth δ_o ($\sim c/\omega_p$) correspondingly smaller. The cyclotron radius R_c is larger than for Bi in the relevant field range ($R_c \sim 0.2$ μm at ~ 3.5 T). The ratio δ_o/R_c, which can be taken as a rough measure for the degree of locality, therefore is at least an order of magnitude smaller as compared to Bi.

Relatively little experimental work has been done on investigating the electrodynamical properties of Sb. The Azbel-Kaner cyclotron resonance technique at microwave frequency has been employed to investigate the Fermi surface long ago[85]. Similar experiments, performed at FIR frequencies were reported in a conference abstract only[86]. Alfvèn wave propagation has been studied at low frequencies (< 50 GHz) and high fields only recently in the Faraday[87] as well as in the Voigt[88] configuration. An experimental investigation of the kind of magnetoplasma effects, presently discussed for Bi (local CR, HR, DA), to our knowledge is not available.

The typical results as obtained from a local calculation analogous to that used for Bi are displayed in Fig. 24 for a frequency and field range relevant to the experiments. For these calculations and those to follow, a density of $5.5 \cdot 10^{19}$ cm^{-3} [84] was assumed, mass parameters were taken from Rao et al.[89] (with electrons and holes interchanged in accordance with later interpretations), while the lattice dielectric constant ($\varepsilon_1 \sim 80$ [90]) is not relevant at these frequencies. The calculations apply to the experimental configurations: Voigt geometry, with the field along the two crystal symmetry axes in the trigonal plane . The spectra obtained are more complicated than the corresponding ones for Bi. With B along the binary axis, two peaks are obtained in the E//B polarization. These correspond to the dielectric anomalies associated with cyclotron resonance of holes (lowest field peak) and electrons, both of which have tilted orbits in this case. The plasma consists now of 4 distinct types of carriers: B, C-electrons and holes and the A-electrons and holes. According to Smith et al.[7], there should be 4-1=3 hybrid resonances for E ⊥ B which occur in Fig. 24 (B//bin) around 1.6 , 3 and 6 T. Between the hybrids, DA's occur. (The small dip around 6.5 T should be ignored for the present discussion. It corresponds to the heavy hole CR, but its physical origin will not be discussed. It is smeared out in the finite τ calculations applicable to the experiments). The general character of the spectrum for B//bis is similar to B//bin, but mode mixing is very strong as can be seen in Fig. 24.

The experimental results obtained for B//bin are given in Fig. 25. Some calculated curves are also given, but these should be regarded as a guide to the eye, rather than a detailed comparison, for reasons to be discussed. The most notable difference between the experimental curves and the local calculations is the magnitude of the signal. The observed change in the surface impedance (or absorption) are in the order of 10%, which have hardly a relation to the calculations, which give absorption changes of more than an order of magnitude. The calculated absolute peak absorption is in the order of 1% only, while the measured absolute overall absorption (e.g. at zero field) is certainly at least as high as 1% to our estimate (although measurements of the absolute absorption values are very difficult to make). If it were only for these facts, a serious comparison of experiment with the calculations would not make much sense, because of the arbitrariness of the scaling.

The point to be noted however is the remarkable correspondence of the pronounced structure observed in the absorption with the local magnetoplasma

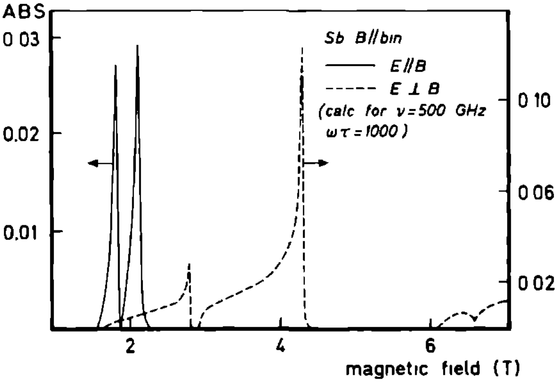


Fig. 24a. Theoretical absorption behaviour of Sb due to plasma resonances and dielectric anomalies according to the classical (local) theory (ref. 7). The curves were drawn for ν=250 GHz and ωτ=1000. Voigt configuration, magnetic field parallel to binary axis. (Compare with Fig. 2 for the Bi case.)

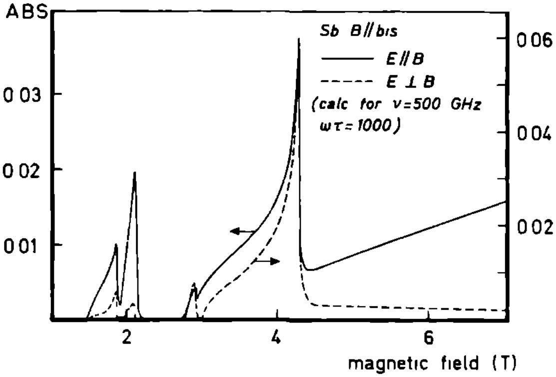


Fig. 24b. As Fig. 24a but with B along a bisectrix axis.

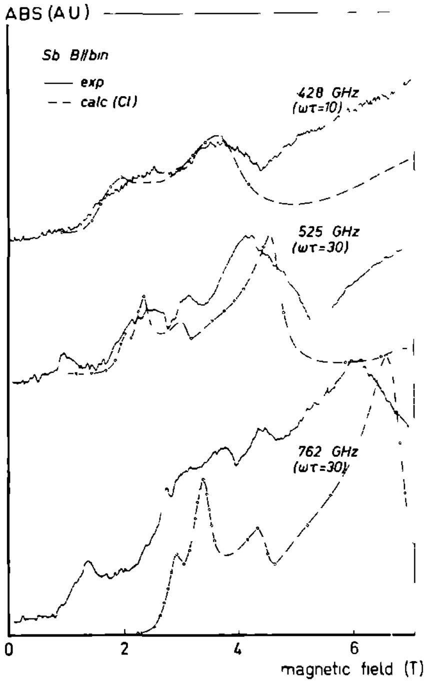


Fig. 25. Absorption as a function of magnetic field B for Sb with B along a binary axis at three different frequencies. Experimental results and calculations. For ωτ-values: see text.

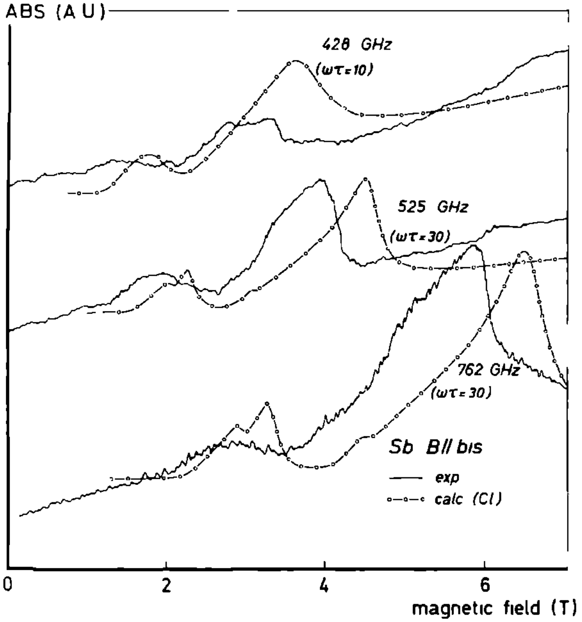


Fig. 26. As Fig.25, but with B along a bisectrix axis.

effects. At 428 GHz the rise in absorption around 1.5 T corresponds to the (unresolved) cyclotron resonances of electrons and holes, while the absorption decrease around 4 T corresponds to the second DA for E ⊥ B (see Fig. 24). The structure associated with the first DA for E ⊥ B is poorly resolved. The same behaviour is seen at 525 GHz, with a somewhat higher resolution, and also a more clear structure around the first DA for E ⊥ B (∿ 3 T). At 762 GHz, the cyclotron resonances of electrons and holes (∿ 2.8 T) are even somewhat resolved. The most deviating feature in Fig. 25 seems to be the absorption rise starting immediately beyond the second DA for E ⊥ B (∿ 4.2 T at 428 GHz and ∿ 5.2 T at 525 GHz). It turns out that the uncertainty in the mass parameters might account for the apparent discrepancy. In the literature, the data regarding the mass parameters scatter rather much; only the cyclotron resonance of light electrons and holes are reasonably consistent. The scatter might be related to the deviations from ellipsoidity of the carrier pockets which give apparently different values in different experiments, when interpreted within an ellipsoidal model. A discrepancy with literature data regarding the mass parameters also were found in the Alfvèn wave studies of refs. 87 and 88. When using e.g. the mass parameters from Mori et al.[91] and referring to Fig. 24 (B//bin), the hybrid resonance at 6.1 T moves to 5.1 T, leaving the rest of the spectrum reasonably unchanged. Such a lower field position of the last hybrid would be much more consistent with our data (see Fig. 25). The mass parameters of Mori et al.[91] give however the same cyclotron mass for the light electrons and holes, which resonances are resolved in the data at 762 GHz.

The structure in the data occurring at low fields, clearly seen at 525 and 762 GHz, which apparently is not accounted for by the local theory, can be attributed to subharmonic (Azbel-Kaner) cyclotron resonances. They are of comparable magnitude to the other absorption structures, which is in sharp contrast to the subharmonics observed in Bi.

In the calculated curves of Fig. 25, rather arbitrary values for τ were chosen. At 428 GHz a low value was used to simulate the experimental rather smooth behaviour. At the other two frequencies, a higher τ had to be used to resolve some of the structure also present in the data, e.g. the dip between the CR's at 762 GHz. On the other hand, the rather sharpness of some of the dips present in the data suggests that the actual relaxation time is still higher. From the literature on Sb it turns out that in order to obtain high quality surfaces and low relaxation rates, utmost care is required in handling the sample, whereas repeated thermal cycling should be avoided. It was found in the present work that thermal recycling did not significantly influence the spectra, which did neither change much after the sample had been annealed. This indicates that the relaxation time of the sample is lower than could have been obtained in principle.

The data for B along the bisectrix direction are shown in Fig. 26. They need not much more discussion beyond the general discussion given for the binary axis data. The spectra are less structured than those for the binary orientation but this is in accordance with the calculations. The discrepancy between measured and calculated position should not be considered very significant, in view of the already mentioned uncertainty in the mass parameters.

From the comparison of the experimental data obtained for Sb with the results from a local theory, it is evident that the conditions are highly nonlocal. This is also born out by the relatively strong subharmonic resonances. It is therefore interesting that nevertheless the structure in the surface impedance can be unambiguously correlated with local magnetoplasma effects such as hybrid resonances and dielectric anomalies.

Influence of nonlocal effects were also observed in the Alfvèn wave propagation studies[87, 88]. Because the mass density is high, the wave velocity is relatively low and becomes comparable to, or even smaller than, the Fermi velocity at low fields. This leads to important collisionless damping effects and distinct onset of wave propagation at a certain (high) field value, which might have been observed in refs. 87 and 88. Analogous effects have been observed for Bi in corresponding field and frequency ranges[65, 66, 6]. Besides distinct onset of wave propagation due to spatial dispersion, deviations from the classical dispersion relation were observed[88] and also attributed to nonlocal effects. Nonlocal effects observed in the Alfvèn wave propagation do not seem to be much more important or qualitatively different however from the nonlocal influences in Bi in corresponding field and frequency ranges[59,60]. It should be noted however that the data from refs. 87 and 88 wave obtained in high field only, i.e. for the Voigt configuration well above the highest hybrid, so outside the field range used for the present experiments (Figs. 25 and 26).

A nonlocal theory, which is relevant to the case of Sb also under the conditions of our experiment, exists only in the recent work of Saermark et al.[63], already mentioned in connection with Bi, for which some explicit results were evaluated in ref. 63. In a nonlocal theory, the wave dispersion consists in principle out of infinitely many branches, so that for a given field value there are correspondingly many values of q. Concepts as hybrid resonance and dielectric anomalies, which are in the local theory defined as

$q = \pm \infty$, respectively $q = 0$, are no longer precisely defined. It was shown[63] however, that the dispersion relations may exhibit a peculiar behaviour near field positions where local HR's or DA's exist. Anomalies in the dispersion relation directly influence the surface impedance also. Near a local hybrid resonance for instance, two modes of the dispersion relation cross and hybridize and such anomaly will be reflected in the absorption . Such hybridization generally occurs over an extended magnetic field range instead of a well defined field value[63].

Apart from the magnitudes of the signals already discussed, the lineshapes strongly deviate from the local calculations and cannot be reconciled with it by adjusting relaxation times. In this connection it is interesting to remark that the heavy hole cyclotron resonance would be expected near a field value where the high field slope of the second DA for $E \perp B$ is located (∿ 4.5 T at 525 GHz), if the mass parameters of Mori et al.[91] were used . (With the mass parameters of Rao et al.[89] this resonance is expected at much higher fields (6.8 T at 525 GHz)).

It would be highly interesting to subject the results of Figs. 25 and 26 to a detailed comparison with a nonlocal theory. Such a detailed theory is presently not available and beyond the scope of this work.

5. Conclusion.

To conclude, some of the main results obtained in this work are summarized. In discussing the magnetoplasma effects, which have been investigated in Bi over a large frequency range, it turns out to be natural and clarifying to separate the observed features in essentially collective properties of the plasma (HR's and DA's) and properties which can be understood on a single particle basis (CR). As has been pointed out before however, such a clear distinction needs not to be made from a theoretical point of view. To analyze the results, use has been made of a quantummechanical expression for the conductivity, which depends on theoretical expressions for the velocity matrix element and takes fully into account the (nonparabolic) energy dispersion of the carriers. However possible special aspects related to purely quantum theoretical concepts as electron spin have been neglected. Also ignored is the spatial variation of the fields compared to the relevant physical length R_c (local limit).

At high fields and frequencies, the cyclotron resonances show pronounced structure which is rich in details. The most remarkable observation is the distinct effect which $k \neq 0$ contributions can have and which are a consequence of the nonparabolic band. Previous far-infrared cyclotron resonance (FIRCR) investigations of the nonparabolic band of Bi exist[19], but their analysis was based on $k = 0$ transitions only. It is common practice to derive band structure parameters from FIRCR based upon an interpretation which assumes $k = 0$ transitions only. Examples, relevant to the present work include the nonparabolic degenerate semiconductors of the PbTe type[15] and the nonparabolic semimetal pyrolytic graphite[92]. The present work does not invalidate such an analysis; it is however shown that $k \neq 0$ contributions may lead to distinct fine structure on the fundamental CR line shape in degenerate materials, which may be used to obtain additional data on the band parameters. Moreover, a $k \neq 0$ analysis is essential in those exceptional cases where no $k = 0$ transitions are possible, such as our data at 2310 GHz (Fig. 13) or the resonances occurring just above the quantum limit B_{qu}. In addition, it was found in the case of Bi that $k = 0$ transitions are usually dominant, not merely by their relatively high joint density of states, but also because of nonlocal effects. The complete k-dependence of the transition energies is usually taken into account in detailed CR line shape studies of semiconductors. An example of such case is the prototype nonparabolic semiconductor InSb[93], where the $k \neq 0$ contributions lead to a clearly identifyable assymmetric line broadening[94]. Distinct structure has however not been observed, because of the very low Fermi level in these materials, making the material practically (or completely) nondegenerate.

A very prominent feature in our data is the so called limiting resonance, which has been observed over a broad frequency range in the characteristically different regions below, as well as above the plasma frequency. To our knowledge a distinct structure in the quantum cyclotron resonance at the limiting k-values has not been reported earlier, neither in Bi nor in another material . Its usefulness for detailed band structure investigations is evident and it is expected to depend on such details as deviations of nonellipsoidity. As for Bi the band structure parameters are supposed to be known sufficiently well from the literature, the limiting resonance has not been exploited in the present work to extract band structure information from it and neither was the experiment adapted for this purpose (as regards alignment accuracies). The question concerning the discrepancy between measured and calculated positions must therefore be left open.

Apart from the CR's, the high-frequency magnetoplasma properties can be described surprisingly well using a simple classical model even when it is in a highly quantized state, provided the FL-modulation is properly taken into account. The FL-modulation itself is a quantum effect and is introduced in the classical theory in a phenomenological way only. The validity of the classical description even in quantizing fields, seems nontrivial, because the quantum model used throughout this work leads to distinctly different results.

In the deviating field region, $B > B_{qu}$, the matrix element as it appears in the quantum calculations, plays an important role. In the present work, the well accepted expression[49] for the matrix element in the two band model, which was however u·ed in a reasonable and well defined approximation, has been applied as a reference in the theoretical calculations. When using a more simple expression[51], the results of the quantum calculation reduce practically to the (quasi-) classical calculations (apart from CR-features), in excellent agreement with the data, over the complete frequency range.

With the alternative matrix element[51], line shapes are also nearly equal in the classical and quantum model in such unique cases where a DA coincides with B_{qu}. The suggestion made in connection with Fig. 6, that the experimental lineshape of the DA reflects a direct influence of the two band model, therefore seems to loose part of its validity. It is of course interesting whether such exceptional conditions as in Fig. 6 may revail details of the matrix element not contained in any of the two expressions discussed here.

The average cyclotron frequency (or energy difference between Landau levels) in the two band model is strongly field-(or frequency-) dependent in the quantum limit. This effect would lead to a substantial different behaviour from that expected for a parabolic band. The multiplication factor for the parabolic band matrix element, to obtain the nonparabolic band matrix element (see eq. (16) or (18)) acts therefore to cancel the field dependence of the average cyclotron frequency. The factor (18c) serves this purpose as can easily seen in case $\omega = 0$ and $\tau = \infty$; then the following identity is easily verified[51]:

$$\frac{1}{(E_\nu - E_{\nu'})^2} \cdot \frac{(E_g)^2}{(E_\nu + E_{\nu'})^2} = \frac{1}{(E^P_\nu - E^P_{\nu'})^2} \qquad (23)$$

where E_ν denote energy values for the nonparabolic band, given by eq.(9) and E^P_ν the energy values of the parabolic band, eq.(15). The left hand side of (23) is the resonant part of eq. (13) times the multiplication factor of the matrix element. Therefore the quantum expression for the conductivity (13) for a nonparabolic band of the two band model is identically the same to that for a parabolic band . It is evidently nearly equal for finite ω

and τ if $\hbar\omega$, $\frac{\hbar}{\tau} << E_\nu - E_{\nu'}$,which is clearly fulfilled at fields where the deviating dielectric anomalies occur in the present work. It is not significant that the right hand side of eq. (13) seems to contain the band bottom mass instead of the Fermi level mass, because the matrix element contains another mass factor, so that the constant of proportionality between m and m_F cancels. It should be noted that in the local approximation the matrix element has no explicit frequency-(or q-) dependence and is the same to that appearing in dc galvanomagnetic transport problems. As has been pointed out by Hansen[51], the equivalence of the quantum conductivity expression in the parabolic and nonparabolic band is required also to make such fundamental concepts as the recently discovered quantized Hall resistance[95] in two-dimensional semiconductor space charge layers independent of deviations from parabolicity, within the framework of simple transport theories.

A direct and useful result from the present work is that the collective high-frequency magnetoplasma properties can be largely analyzed using a completely classical treatment even when the plasma is in a highly quantized state. (Possible charge carrier oscillations need to be taken into account however). Such a classical behaviour might have fundamental reasons, but is presently only justified for the specific case of Bi, with a specific choice for the velocity matrix element.

It has been shown that nonlocal effects do clearly play a role in the case of Bi but not in an essential way. This is clearly not true in the case of Sb, as expected, where nonlocal effects drastically modify the local behaviour. The subharmonic resonances in Bi have not been discussed in detail, but they were referred to in passing as nonlocal (Azbel-Kaner-like) effects, following the general belief. It would be of interest however, to investigate to what extend they could appear in the local theory, as a result of known[4] small deviations from ellipsoidity of the FS. Subharmonics of that kind were observed for the semimetal pyrolytic graphite[92] and could be recognized as such because the selection rule for the allowed subharmonic number is consistent with the orbital symmetry (the relevant orbits in graphite are strongly trigonal warped).

As a final point, we remark that the frequency dependence of the relaxation time which was initially chosen, has sufficed rather well over the whole frequency range in both orientations. A frequency-independent relaxation time would have given clearly too high relaxation times in the higher frequency range.

Acknowledgements.

We are most grateful to Ir. L.W.M. Schreurs for providing, cutting and orienting the samples and to Drs. M.J.H. van de Steeg for the use of his plot-subroutines.

REFERENCES

1. V.S. Edelman, Adv. Phys. 25, 555 (1976).
2. B. Lax and J.G. Mavroides in *Advances in Solid State Physics*, edited by F. Seitz and D. Turnbull, (Academic, New York, 1960), Vol. 11.
3. V.S. Edelman and M.S. Khaikin, Zh. Eksp. Teor. Fiz. 49, 107 (1965) [Sov. Phys. - JETP 22, 77 (1966)].
4. R.J. Dinger and A.W. Lawson, Phys. Rev. B 7, 5215 (1973).
5. R.T. Isaacson and G.A. Williams, Phys. Rev. 177, 738 (1969).
6. V.S. Edelman, Usp. Fiz. Nauk 102, 55 (1970) [Soviet Phys. Uspekhi 13, 583 (1971)].
7. G.E. Smith, L.C. Hebel and S.J. Buchsbaum, Phys. Rev. 129, 154 (1963).
8. E.D. Palik and J.K. Furdyna, Rep. Prog. Phys. 33, 1193 (1970).
9. S. Perkowitz, Phys. Rev. 182, 828 (1969).
10. V.S. Edelman, Zh. Eksp. Teor. Fiz., Pis'ma Red. 9, 302 (1969) [JETP Lett. 9, 177 (1969)].
11. V.S. Edelman, A.S. Garevskii and V.Ya Demikhovskii, Fiz. Tverd. Tela 16, 3739 (1974) [Sov. Phys. Sol. State 16, 2435 (1975)].
12. W. Braune, J. Lebech and K. Saermark, Phys. Stat. Sol. (b) 87, 527 (1978).
13. N.B. Brandt and S.M. Chudinov, J.Low Temp. Phys. 8, 339 (1972).
14. H.R. Verdún and H.D. Drew, Phys. Rev. B 15, 5636 (1977).
15. S. Baldwin and H.D. Drew, Phys. Rev. Lett. 45, 2063 (1980).
16. S.W. Mcknight and H.D. Drew, Phys. Rev. B 21, 3447 (1980).
17. W.S. Boyle and A.D. Brailsford, Phys. Rev. 120, 1943 (1960).
18. U. Strom, H.D. Drew and J.F. Koch, Phys. Rev. Lett. 26, 1110 (1971).
19. U. Strom, A. Kamgar and J.F. Koch, Phys. Rev. B 7, 2435 (1973).
20. R.L. Blewitt and A.J. Sievers, J. Low Temp. Phys. 13, 617 (1973).
21. H.R. Verdún and H.D. Drew, Phys. Rev. Lett. 33, 1608 (1974).
22. H.R. Verdún and H.D. Drew, Phys. Rev. B 14, 1370 (1976).
23. H.D. Drew and U. Strom, Phys. Rev. Lett. 25, 1755 (1970).
24. A. Kamgar and H. Schaber, J. Low Temp. Phys. 26, 423 (1977).
25. J.C. Burgiel and L.C. Hebel, Phys. Rev. 140, A925 (1965).
26. D.L. Carter and J.C. Picard, Solid State Commun. 5, 719 (1967).
27. R.N. Brown, J.G. Mavroides and B. Lax, Phys. Rev. 129, 2055 (1963).
28. M. Maltz and M.S. Dresselhaus, Phys. Rev. B 2, 2877 (1970).
29. M.P. Vecchi and M.S. Dresselhaus, Phys. Rev. B 9, 3257 (1974).
30. M.P. Vecchi, J.R. Pereira and M.S. Dresselhaus, Phys. Rev. B 14, 298 (1976).

31. V.V. Druzhinin, A.I. Pavlovskii, O.M. Tatsenko, M.I. Dolotenko, A.I. Bykov and N.P. Kolokol'chikov, Pis'ma Zh. Eksp. Teor. Fiz. 32, 523 (1980) [JETP Lett. 32, 502 (1980)].

32. S. Takano and H. Kawamura, J. Phys. Soc. Japan 28, 348 (1970).

33. V.S. Edelman, Zh. Eksp. Teor. Fiz. 68, 257 (1975) [Sov. Phys.-JETP 41, 125 (1975)].

34. F.P. Missell and M.S. Dresselhaus, Phys. Rev. B 5, 1364 (1972).

35. M. von Ortenberg in *Infrared and Millimeter Waves*, vol. 3, K.J. Button, ed. (Academic, New York, 1980), Ch. 6.

36. S.J. Allen, Jr., L.W. Rupp, Jr., and P.H. Schmidt, Phys. Rev. B 7, 5121 (1973).

37. P. Goy and B. Castaing, Phys. Rev. B 7, 4409 (1973).

38. K. Saermark, J. Lebech and W. Braune, J. Phys. F. : Metal Phys. 11, 1691 (1981).

39. J.M. Peech, Technical Report No 9, Division of Engineering and Applied Physics, Harvard University, Cambridge, Massachusetts (USA), (september 1974).

40. M.Ya. Azbel' and E.A. Kaner, J. Phys. Chem. Solids 6, 113 (1958).

41. B. Lax, K.J. Button, H.J. Zeiger and L.M. Roth, Phys. Rev. 102, 715 (1956).

42. A.P. Cracknell, Adv. Phys. 18, 681 (1969).

43. M.H. Cohen and E.I. Blount, Philos. Mag. 5, 115 (1960).

44. G.E. Smith, G.A. Baraff and J.M. Rowell, Phys. Rev. 135, A1118 (1964).

45. G.A. Baraff, Phys. Rev. 137, A842 (1965).

46. R. Stegmann and G. Oelgart, Phys. Stat. Sol. (b) 87, 485 (1978).

47. J.J. Quinn and S. Rodriguez, Phys. Phys. Rev. 128, 2487 (1962).

48. J.J. Quinn, Phys. Rev. 135, A181 (1964).

49. P.A. Wolff, J. Phys. Chem. Solids 25, 1057 (1964).

50. A. Miklavc and H.D. Drew, Phys. Rev. B 23, 5269 (1981).

51. O.P. Hansen, J. Phys. C: Solid State Phys. 14, 5501 (1981).

52. V.K. Arora and M. Jaafarian, Phys. Rev. B 13, 4457 (1976).

53. S. Tosima, J.J. Quinn and M.A. Lampert, Phys. Rev. 137, A883 (1965); M.P. Greene, H.J. Lee, J.J. Quinn and S. Rodriguez, Phys. Rev. 177, 1019 (1969).

54. H.R. Drew, Phys. Rev. B 5, 360 (1970).

55. B.E. Meierovich, Zh. Eksp. Teor. Fiz. 58, 1412 (1970) [Sov. Phys. - JETP 31, 756 (1970)].

56. J. Nakahara, H. Kawamura and Y. Sawada, Phys. Rev. B 3, 3155 (1971).

57. M. Giura, R. Marcon and P. Marietti, Phys. Rev. B 21, 4419 (1980); M. Giura and R. Marcon, Phys. Rev. B 23, 676 (1981); M. Giura, R. Marcon and P. Marietti, Solid State Commun. 40, 659 (1981).

58. V.L. Gurevich, V.G. Skobov and Yu. A. Firsov, Zh.Eksp. Teor. Fiz. 40, 786 (1961) [Sov. Phys. - JETP 13, 552 (1961)].

59. W. Braune, W. Schneider, J. Lebech and K. Saermark, Phys. Stat. Sol. (b) 93, 269 (1979).

60. W. Braune, J. Lebech and K. Saermark, J. Phys. F: Metal Phys. 9, 223 (1979).

61. J. Lebech, K. Saermark and W. Braune, Phys. Lett. 75^{A}, 417 (1980).

62. K. Saermark, J. Lebech and H. Johansen, Solid State Commun. 32, 1163 (1979).

63. K. Saermark, J. Lebech, H. Johansen and W. Braune, J. Phys. F: Metal Phys. 10, 1443 (1980).

64. I. Yokota, J. Phys. Soc. Japan 21, 1851 (1966).

65. W.L. Lupatkin and C.A. Nanney, Phys. Rev. Lett. 20, 212 (1968).

66. C. Guthman, J.P. D'Haenens and A. Libchaber, Phys. Rev. B 4, 1538 (1971).

67. V.S. Edelman and S.M. Cheremisin, Zh. Eksp. Teor. Fiz. Pis'ma Red. 11, 373 (1970) [JETP Lett. 11, 250 (1970)]; S.M. Cheremisin, V.S. Edelman and M.S. Khaikin, Zh. Eksp. Teor. Fiz. 61, 1112 (1971) [Sov. Phys. - JETP 34, 594 (1972)].

68. R.T. Isaacson and G.A. Williams, Phys. Rev. Lett. 32, 26 (1969).

69. N.B. Brandt, V.A. Yastrabova and Ya. G. Ponomarev, Fiz. Tverd. Tela 16, 102 (1974) [Sov. Phys. - Solid State 16, 59 (1974)].

70. T. Akahane and Y. Uemura, J. Phys. Soc. Japan 31, 990 (1971).

71. W. Braune, J. Lebech and K. Saermark, Phys. Stat. Sol. (b) 97, 247 (1980).

72. M.S. Dresselhaus and J.G. Mavroides, Solid State Commun. 2, 297 (1964).

73. S. Mase, J. Phys. Soc. Japan 21, 243 (1966).

74. D.L. Mitchell, E.D. Palik and R.F. Wallis, Phys. Rev. Lett. 14, 827 (1965); D.L. Mitchell, E.D. Palik and R.F. Wallis, J. Phys. Soc. Japan, Suppl. 21, 197 (1966).

75. R.B. Dennis, S.D. Smith and C.J. Summers, Proc. Roy. Soc. Lond. A 321, 303 (1971).

76. M. Lutz, E. Bangert, T. Mizrah and H. Stolze, Phys. Stat. Sol. (b) 71, 523 (1975).

77. M.J. Aubin and A. Rambo, Solid State Commun. 30, 297 (1979).

78. A.S. Saleh and H.Y. Fan, Phys. Rev. B 5, 3972 (1972).

79. E.R. Vol'skii, Zh. Eksp. Teor. Fiz. 46, 123 (1964) [Sov. Phys. - JETP 19, 89 (1964)].

80. M.Ya. Azbel', J. Phys. Chem. Solids 7, 105 (1958).

81. F.W. Spong and A.F. Kip, Phys. Rev. 137, A 431 (1965).

82. G. Abstreiter, J.P. Kotthaus, J.F. Koch and G. Dorda, Phys. Rev. B 14, 2480 (1976).

83. L.C. Hebel and P.A. Wolff, Phys. Rev. Lett. 11, 368 (1963).

84. L.R. Windmiller, Phys. Rev. 149, 472 (1966).

85. W.R. Datars and J. Vanderkooy, IBM J. Res. Develop. 8, 247 (1964).

86. S.J. Allen, Jr., Bull. Amer. Phys. Soc. 15, 1355 (1970).

87. H. Suematsu and S. Tanuma, J. Phys. Soc. Japan 41, 496 (1976).

88. W. Braune, R. Kuhl, J. Lebech and K. Saermark, Solid State Commun. 25, 105 (1978).

89. G.N. Rao, N.H. Zebouni, C.G. Grenier and J.M. Reynolds, Phys. Rev. 133, A 141 (1964).

90. C. Nanney, Phys. Rev. 129, 109 (1963).

91. H. Mori and S. Mase, J. Phys. Soc. Japan 31, 738 (1971).

92. R.E. Doezema, W.R. Datars, H. Schaber and A. Van Schyndel, Phys. Rev. B 19, 4224 (1979).

93. R. Kaplan, B.D. McCombe and R.J. Wagner, Solid State Commun. 12, 967 (1973).

94. O. Matsuda and E. Otsuka, J. Phys. Chem. Solids 40, 809 (1979).

95. K. von Klitzing, G. Dorda and M. Pepper, Phys. Rev. Lett. 45, 494 (1980).

CHAPTER IV. CLASSICAL RECTIFICATION.

IV 1.

A new mechanism for high-frequency rectification at low temperatures in point contacts between identical metals

R W van der Heijden, A G M Jansen, J H M Stoelinga, H M Swartjes, and P Wyder
Research Institute for Materials, University of Nijmegen Toernooiveld, 6525 ED Nijmegen The Netherlands

(Received 6 March 1980, accepted for publication 17 April 1980)

Experimental results are reported of high-frequency (~ THz) radiation detection by metal-metal point contacts at low temperatures as a function of bias voltage The dominant detection mechanism can be attributed to rectification due to electron-phonon-scattering-induced nonlinearity of the I-V characteristics, a process not observed before

PACS numbers 73 40 Ei

Point contacts between two metals have found widespread use during the last decade in quantum-electronic applications Because of their high-speed capabilities, absolute and precise laser-frequency measurements can be made in the region up to the visible This has led to considerable progress in laser spectroscopy[1] and to a determination of the speed of light two orders of magnitude more accurate.[2] Recently, these devices have proved to be of great use in studying fundamental processes in the metals composing the junctions.[3-5]

Despite its great practical use, however, the physical mechanism behind the high-frequency operation of the diode has not yet been clarified satisfactorily It is generally believed that high-frequency currents are flowing through the contact, which has a nonlinear *I-V* characteristic and thereby causes rectification. The origin of the nonlinearity is invariably attributed to a nonlinear tunneling mechanism through the interface. Experimentally, good agreement with such a theory is found in the case of metals having different work functions, i.e., nonidentical electrode materials.[6-10] The case of identical electrodes is more complicated, because here a tunneling model gives nearly linear *I-V* characteristics.[11] Therefore, one often invokes accidental local work-function differences to explain the excellent performance of, e.g., Au-Au diodes.[6] Rigorous verification of such nonlinear tunneling processes, however, is tedious, because the *I-V* characteristics, experimentally as well as theoretically, depend strongly on the detailed nature of the barrier.

Recently, it was shown theoretically[12] that the *I-V* characteristic becomes nonlinear when the extremely different geometry of both electrodes is taken into account, along with their asymmetric heating when exposed to high-intensity laser fields (thermally enhanced field emission). This however, seems to be effective for high-power (~ 100 W) laser beams only.[12]

In this letter, we present new data taken on Cu-Cu point contacts at low temperature which strongly support the idea of a pure rectification process The frequencies used are low enough to neglect photon absorption The nonlinearity causing the rectification in this case, however, is not due to a nonlinear tunneling process, but to an inelastic electron-phonon scattering mechanism This nonlinearity has been studied extensively experimentally,[3 4] a firm theoretical basis also exists which readily explains most of the observed features.[13 14]

The experiments were performed at submillimeter wavelengths, using an optically pumped far-infrared laser as a source. The radiation was guided to the junction using a light pipe ending in a Cu cone with 4-mm final diameter and terminated by a reflecting Cu plate. The junction wire was mounted at the end of the cone, across its diameter. The actual contact was located at the side of the cone. This assembly was immersed in a liquid-helium bath which could be pumped down to 1.5 K. Contacts could be made and adjusted at liquid-helium temperatures by means of a differential screw mechanism. The whisker was made from 50-μm-diam Cu wire electrolytically etched to a sharp point. The post consisted of a piece of Cu etched in HNO_3 but not further treated. *I-V* characteristics as well as first and second derivatives as functions of the applied voltage were recorded using current modulation and phase-sensitive detection. For laser detection, the beam was chopped at 560 Hz and the signal across the junction phase sensitivity detected and recorded as a function of bias voltage Most data have been taken at 525 GHz [571-μm line, CH_3OH (Ref. 15), but performance was checked at 2.52 THz and 246 GHz (118- and 1217-μm line, CH_3OH, respectively[15]).

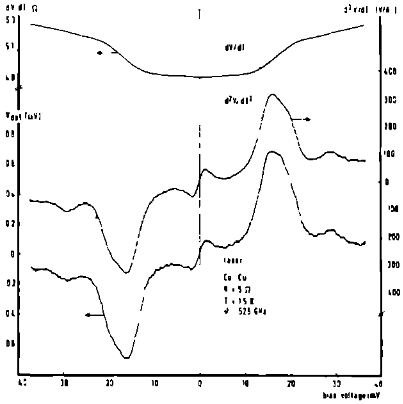


FIG 1 Upper two traces dV/dI and d^2V/dI^2, respectively, of the *I-V* characteristic as a function of dc bias voltage for a Cu-Cu contact Lower trace laser-detected signal as a function of bias voltage Junction resistance 5 Ω, temperature 1 5 K, laser frequency 525 GHz

Figure 1 shows a typical result of the measurements. The upper two traces show the first and second derivatives of the *I-V* characteristic The lower trace is a recording of the laser detection output The power level from the laser is estimated to be between 1 and 10 mW. The first two traces are taken simultaneously, but trace 3 was taken in a separate run for practical reasons. It was checked, however, that the contact had not changed noticeably after taking trace 3 by recording trace 1 and 2 again. Values for the nonlinearity $(d^2I/dV^2)/(dI/dV)$ on the order of 10 V^{-1} are typical, which compare very well with the best values reported for room-temperature devices.[6 8]

Comparing the two lower traces in Fig. 1, it is immediately seen that they are practically identical. This is exactly what is expected for a high-frequency rectification process at a low power level, for then the signal is proportional to $(d^2I/dV^2)/(dI/dV)$.

It has been shown[3 4] that the pronounced nonlinearities at low temperatures of the dc *I-V* characteristic for such contacts could be explained in terms of the electron-phonon interaction The investigated point contacts are in the clean limit, i.e., the electronic mean free path is larger than or comparable to the contact diameter. For this type of contacts, the electrons are accelerated in the electric field caused by the applied voltage and a relaxation of the accelerated electrons via elementary excitations (i.e., phonons) in the metal takes place. These inelastic scattering processes yield a nonlinear current-voltage characteristic The experiment shows that scattering processes with phonons are important; the broadened peak in the d^2V/dI^2 spectrum at 17 meV in Fig 1 corresponds to the two transverse phonon frequencies in Cu, while the peak at 28 meV agrees well with the longitudinal phonon frequency

In a theoretical analysis of the point-contact problem the current through the contact has been calculated by an iterative solution of the Boltzmann equation with the appropriate boundary conditions [13] [14] In the zeroth-order approximation, which ignores inelastic scattering, the field emission current is found to be linear in the applied voltage In next order, the scattering term in the Boltzmann equation is taken into account, which yields a negative correction to the current because the emitted electrons can flow back through the orifice after an inelastic collision, i e , a spontaneous emission of a phonon Neglecting transport efficiencies one finds as the final result in point contact spectroscopy that the measured d^2V/dI^2 curve will be proportional to the electron phonon interaction function $\alpha^2F(\omega)$, $\alpha^2F(\omega)$ is the product of the phonon density of states and the squared matrix element for the electron-phonon interaction, averaged over the Fermi surface The fact that the measured spectrum doesn't vanish above the Debye energy (~ 30 meV for Cu), is ascribed to the nonequilibrium distribution of the phonons, arising from the phonon emission near the contact [16] Stimulated emission and absorption of phonons now give a smooth "background" signal which is nonzero above the Debye energy

A zero bias anomaly is typical for this kind of junction[3] [4] and is seen as a resistance maximum, minimum, or no structure at all in an irreproducible way Several mechanisms have been suggested for this anomaly in the case of tunnel junctions,[17] while recently a resistance maximum could be attributed to the Kondo effect in Cu Cu and Au-Au point contacts, when deliberately doped with magnetic impurities [18]

In the theoretical analysis of the point-contact spectroscopy, the detailed nature of the interface does not affect the main results derived for a metallic contact If some barrier exists at the interface the results remain qualitatively unchanged, but an energy-independent tunneling probability can be taken into account [14] In fact, for low ohmic junctions ($< 10\,\Omega$) we believe we have real metallic contacts For high ohmic junctions this is no longer reasonable as it would require an unrealistically small contact diameter (< 40 Å)

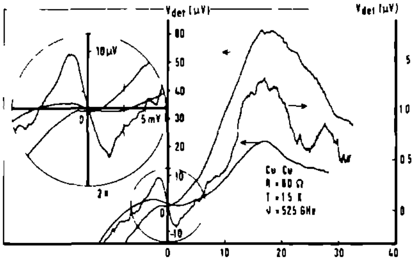


FIG 2 Laser detected signal as a function of bias voltage for different laser powers (uncalibrated) for a Cu Cu contact Inset magnified view of the structure around $V = 0$ A polarity reversal is indicated by the dashed line Junction resistance $\sim 60\,\Omega$ temperature 1 5 K laser frequency 525 GHz

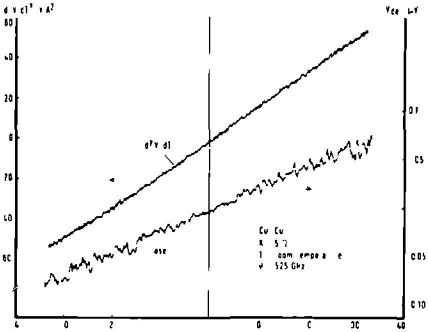


FIG 3 Upper and lower trace d^2V/dI^2 and laser-detected signal as a function of dc bias voltage respectively for a Cu Cu contact Junction resistance 5 Ω T room temperature laser frequency 525 GHz

It is of interest to know what the expected ultimate speed of response for this new low-temperature mechanism will be From the above discussed operation principle it follows that the dc I-V characteristic will be followed at frequencies such that $\omega\tau < 1$, where τ is the (energy dependent) electron-phonon scattering time and ω the angular frequency of the radiation Typical values for τ at electron energies comparable to the phonon energies are 10^{13} sec in metals This value is considerably larger than that usually assumed for room temperature devices It must be noted that low-energy electrons (long τ) do not establish a significant nonlinearity, even at dc First, the phonon density of states is too low and second, long mean free-path electrons have a low backscattering probability (small angle of view[14]) This explains why the high-frequency rectification process closely follows the dc characteristic

Among the problems in assessing the rectification process often encountered in video detection experiments are spurious signals due to thermoelectric and bolometric effects [19] However, we believe that these effects are negligible in the present case with low power levels The excellent correspondence between the directly measured d^2V/dI^2 spectra and the laser-detected signal strongly suggests detection by rectification Even more evidence is given by results on sensitive junctions with a zero-bias anomaly, reversed in sign with respect to the phonon peaks At relatively high power levels, the structure is strongly broadened and smeared out owing to overmodulation (Fig 2), especially nicely seen on the zero-bias anomaly and longitudinal phonon peak, which extend only over a few mV Because of the reverse signs of the zero-bias anomaly and the phonon peak, we note a polarity reversal as a function of power at a bias voltage indicated by the dashed line in Fig 2 Further, the results could not be reproduced by heating the post at a low frequency with a heater coil glued tightly around it

It is interesting to consider possible consequences for room temperature operation of the diode From a detailed study of the temperature dependence,[20] it was shown that the peaks broaden and disappear at higher temperatures, but a background remains up to room temperature. Figure 3 shows a d^2V/dI^2 and laser-induced signal plot taken at room temperature. This plot was taken in the same setup as for Fig. 1, except that the dewar was not filled with liquid helium. The same whisker and post were used as for Fig. 1, but the actual contact was different. Both d^2V/dI^2 and the detected signal are smaller than at liquid-helium temperatures, but are still in correspondence. Although there is no rigorous proof, these results are not inconsistent with an electron-phonon-scattering-induced nonlinearity which causes rectification at room temperature.

In conclusion, we may state that experimental evidence is given for a high-frequency rectification process based upon a nonlinearity in the I-V characteristics caused by inelastic electron-phonon interaction. In a Cu-Cu contact at low temperatures and for low bias voltages, this effect certainly dominates all other possible effects in the submillimeter region. This nonlinearity is a very pertinent effect which occurs always in the same way, not depending on details of the contact, for the whole range of resistance investigated (1 Ω–1 kΩ). The bias voltage for which it reaches a maximum, depends on bulk material properties only. The experimental results can be interpreted on a firm theoretical basis. Finally, the mechanism might be active even at room temperature.

We are most grateful to Dr. H. J. A. Bluyssen and A. F. van Etteger for their invaluable help and advice in operating the laser system, which was built on their design. Part of this work has been supported by the "Stichting voor Fundamenteel Onderzoek der Materie" with financial support from the "Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek."

[1] For comprehensive reviews see A. Javan, in *Fundamental and Applied Laser Physics*, edited by M. S. Feld, A. Javan, and N. A. Kurnit (Wiley, New York, 1973); K. M. Evenson and F. R. Petersen, in *Laser Spectroscopy of Atoms and Molecules*, edited by H. Walther (Springer, Berlin, 1976).

[2] K. M. Evenson, J. S. Wells, F. R. Petersen, B. L. Danielson, G. W. Day, R. L. Barger, and J. L. Hall, Phys. Rev. Lett. **29**, 1346 (1972).

[3] A. G. M. Jansen, F. M. Mueller, and P. Wyder, Phys. Rev. B **16**, 1325 (1977); A. G. M. Jansen, P. Wyder, and F. M. Mueller, Science **199**, 1037 (1978).

[4] I. K. Yanson, Zh. Eksp. Teor. Fiz. **66**, 1035 (1974) [Sov. Phys. JETP **39**, 506 (1974)].

[5] K. C. Liu, C. Davies, Jr., and A. Javan, Phys. Rev. Lett. **43**, 785 (1979).

[6] S. I. Green, P. D. Coleman, and J. R. Baird, presented at the Symposium on Submillimeter Waves, Polytechnic Institute of New York, Brooklyn, 1970 (unpublished), pp. 369–389.

[7] S. P. Kwok, G. I. Haddad, and G. Lobov, J. Appl. Phys. **42**, 554 (1971).

[8] S. M. Faris, T. K. Gustafson, and J. C. Wiesner, IEEE J. Quantum Electron. QE-**9**, 737 (1973).

[9] S. M. Faris, B. Fan, and T. K. Gustafson, Appl. Phys. Lett. **27**, 629 (1975).

[10] S. M. Faris and T. K. Gustafson, Appl. Phys. Lett. **25**, 544 (1974).

[11] C. B. Duke, *Tunneling in Solids* (Academic, New York, 1969).

[12] A. A. Lucas and P. H. Cutler, Solid State Commun. **13**, 361 (1973); T. E. Sullivan, A. A. Lucas, and P. H. Cutler, Appl. Phys. **14**, 289 (1977); N. M. Miskovsky, S. J. Shepherd, P. H. Cutler, T. E. Sullivan, and A. A. Lucas, Appl. Phys. Lett. **35**, 560 (1979).

[13] I. O. Kulik, and A. N. Omel'yanchuk, and R. I. Shekhter, Fiz. Nizk. Temp. **3**, 1543 (1977) [Sov. J. Low Temp. Phys. **3**, 740 (1977)].

[14] A. P. van Gelder, Solid State Commun. **25**, 1097 (1978).

[15] J. J. Gallagher, M. D. Blue, B. Bean, and S. Perkowitz, Infrared Phys. **17**, 43 (1977).

[16] A. P. van Gelder, A. G. M. Jansen, S. Strassler, and P. Wyder, J. Phys. (Paris) **39**, C6–602 (1978); A. P. van Gelder (unpublished).

[17] E. L. Wolf, in *Inelastic Electron Tunneling Spectroscopy*, edited by T. Wolfram (Springer, Berlin, 1978).

[18] A. G. M. Jansen, A. P. van Gelder, P. Wyder, and S. Strassler (unpublished).

[19] S. M. Faris, B. Fan, T. K. Gustafson, and J. P. Heritage, IEEE J. Quantum Electron. QE-**10**, 755 (1974).

[20] A. P. van Gelder, A. G. M. Jansen, and P. Wyder (unpublished).

IV 2. HIGH-FREQUENCY RECTIFICATION DUE TO THE NONLINEAR SPREADING RESISTANCE IN MIM AND McM POINT CONTACT JUNCTIONS AT LOW TEMPERATURE.

ABSTRACT

The high-frequency properties of normal metal point contacts at low temperatures are investigated by observing their video responsivity to far-infrared laser radiation as a function of bias voltage. Detection occurs through rectification, due to the nonlinear current-voltage characteristic of the diode. The nonlinearity is caused by the inelastic scattering of electrons by phonons and thus can be attributed solely to the spreading resistance.

The detection was investigated at temperatures up to room temperature and for junctions of varying resistance. A comparison has been made of the relative merits of this detector with other high-frequency point contact devices.

2. 1. Introduction.

Point contact diodes are by now very well established nonlinear devices for use as detectors and mixers at far-infrared and infrared frequencies. They owe their usefulness to the extreme short response times, allowing absolute laserfrequency measurements to be made in a region approaching the visible light frequencies, which is of great importance for high precision laserspectroscopy and in establishing fundamental length and time standards[1]. Point contact configurations are used for Josephson junctions, Schottky diodes and metal-insulator-metal (MIM) junctions[2]. When exposed to a radiation field, high-frequency currents are generated across the contact by the antenna action of the wire, being part of the contact[3]. If the contact has a nonlinear current-voltage characteristic for the high-frequency signal, the diode can be used for detection by rectification[4,5], harmonic generation and mixing[6] and sideband generation[7].

For the origin of the nonlinearity in case of point contacts between metals, an electron tunneling mechanism through a thin oxide layer is generally assumed[4]. Other mechanisms however have been proposed or observed also at near-infrared or visible light frequencies. Photo-excitation of electrons across the potential barrier was observed in a thin film geometry[8] and the importance of resonant interband transitions was discussed theoretically[9]. The importance of thermally enhanced and field induced tunneling, in combination with the specific geometry of the point contact configuration at high frequencies and power levels, has been emphasized in a series of theoretical papers[10], and may have been observed experimentally[11]. It has been proposed also[12] that surface plasmons may become important at near optical frequencies. In addition, the usual method of treating the nonlinear I(V) characteristics originating from electron tunneling has recently been critisized[13].

In the present work, the far infrared detecting properties of point contacts between two normal metals have been investigated. For the type of contacts selected for the present work, however, current transport clearly is not due to tunneling; instead, constriction type of contacts (McM) have been used predominantly. Yet, the current-voltage characteristic may be nonlinear, as a result of inelastic electron-phonon scattering processes[14], which is particularly pronounced at liquid Helium temperatures. Rectification of far-infrared frequencies, based on this type of nonlinearity, has been established recently[15].

After description of some experimental details, the physical mechanism for the nonlinearity and the principle of high-frequency operation will be briefly recalled. The experimental results will be presented and discussed and finally possible influence of this type of high-frequency rectification on other types of devices (e.g. the MIM junction), or even the use of constriction type contacts for practical applications, will be discussed.

2. 2. Experimental details.

As the present experiments were aimed at investigating the detection mechanism only, no elaborate coupling scheme has been employed. A schematic outline of the set-up is given in Fig. 1. The laser was a home-built conventional type optically pumped far-infrared waveguide laser[16]. The FIR cavity consists of a 1 m long, 25 mm diameter gold coated metal waveguide, terminated by flat metal mirrors. Hole output coupling is employed, because optimal beam quality is not required,as the radiation is further transported by oversized waveguides (lightpipes). The power level from the laser is in the mW range.

The point contact was mounted at the end of a 13 mm diameter stainless steel lightpipe which ended in a small angle copper cone of 4 mm final diameter. The lightpipe system was placed in a standard glass dewar which can be filled with liquid Helium, so that the point contact was directly surrounded by the Helium (see Fig. 1a). The Helium bath can be pumped down to ∿ 1.5 K.

Contacts can be made and adjusted from the top of the dewar using a differential screw assembly in combination with a cantilever mechanism (Fig. 1b). Adjustment accuracy corresponds to 25 μm per dial turn. The post was mounted in a fixed position, usually such that its front face approximately coincided with the cone side-wall. Mounting the post further inside the pipe did not significantly alter the detection performance, consistent with the antenna action of the wire. An S-shaped spring was bent in the whisker, intended to provide somewhat improved mechanical stability.

The optical cone was closed by a reflecting copper plate. At the outer face of this plate, an Allen & Bradley carbon resistor (not shown in Fig. 1b) was mounted, which can be used as a thermometer or to monitor the laser power during the experiments at a location near the contact, while it is heated by stray radiation. The possibility of normalizing the detected signal, using the resistor as a reference, also existed in order to eliminate possible laser fluctuations or drift. The laser, which was always operated in a free running mode, was usually stable enough that normalization was unnecessary and therefore mostly omitted.

Copper was used for the whisker and post material and was of normal, electrical grade quality. The whisker tip was obtained by standard electrolytical etching. The post was usually only etched in HNO_3 and occasionally mechanically polished. Polishing however turned out to be not very important and was

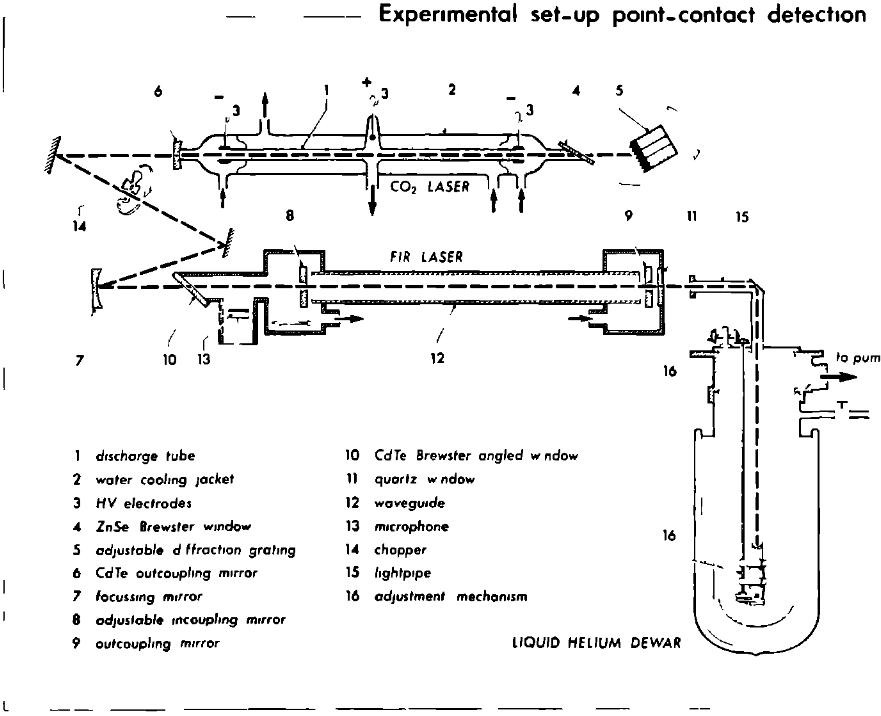


Fig. 1a. Schematic outline of point contact detection experiment, showing the laser and cryogenic dewar.

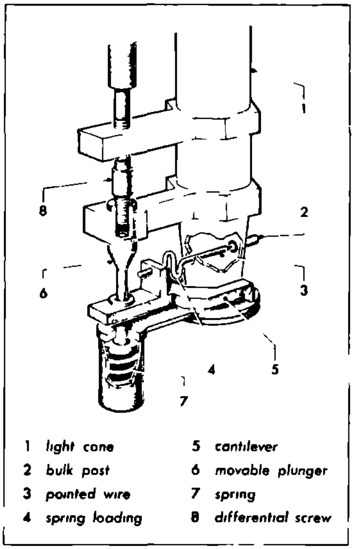


Fig. 1b. Detail of point contact mounting at the lower end of the lightpipe.

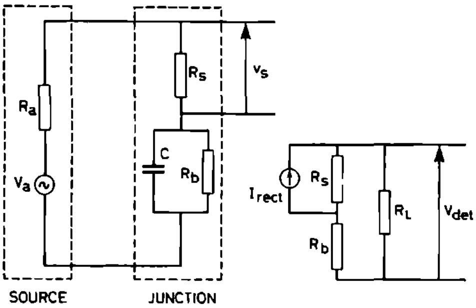


Fig. 2. Equivalent networks for analyzing the point contact detection. Left: ac circuit, right: dc circuit. For video detection $R_L \gg R_b, R_s$, so that it suffices to calculate the open circuit voltage.

usually omitted. Good results were obtained with a surface that looked even somewhat rough to the bare eye.

Current-voltage characteristics of the diode could be recorded directly using a four probe method. First and second derivatives of the I(V) characteristics (dV/dI and d^2V/dI^2) were also measured by modulating the current at low amplitude at 500 Hz, and using phase sensitive detection. The junction was contained in a suitable bridge circuit (see ref. 14).

Laser induced signals were also phase sensitively detected, the CO_2 laser beam being chopped at 560 Hz. To improve sensitivity, a low noise selective preamplifier (PAR 113) was used in front of the lock-in amplifier, optionally preceded by an impedance matching transformer (PAR AM1). Signals were recorded as function of dc bias voltage.

2. 3. Theory.

The high-frequency operation of point contact junctions (MIM, Schottky or even Josephson) is usually analyzed in terms of an equivalent network representation such as given in Fig. 2[5,17-19]. The antenna is represented as a generator with internal impedance R_a and voltage amplitude V_a. This source is loaded by the junction, which in general is assumed to have a barrier resistance R_b shunted by a capacitance C. The resistance R_s is the spreading resistance, which results from the constriction of current flow lines near the contact. As we are only concerned with (∿dc) video detection, the detecting network is trivial; input impedance of the preamplifier can be considered as ∞. Also, the impedance of the dc biasing circuit can be neglected.

2. 3 i. Circuit parameters.

The power which can be maximally delivered by the generator to the load, P_a, is given by $P_a = V_a^2/8R_a$. This power is to be found from antenna theory and described by the antenna capture cross section σ_a. For a plane wave incident at the antenna, σ_a is defined as:

$$\text{power received } (\equiv P_a) = \sigma_a \cdot \text{incident power/unit area} \qquad (1)$$

and can be written as[20]:

$$\sigma_a = (\lambda^2/4\pi)g \qquad (2)$$

where g is the antenna gain and λ the wavelength. g is a strongly directional dependent function, exhibiting the characteristic antenna lobe patterns experimentally observed[3]. For long antennas, its maximum value (for radiation incident along the main lobe) can be approximated by[17]:

$$g_m \sim (120/R_a)\cdot(L/\lambda) \qquad (3)$$

The polarization should be in a plane containing the antenna.

The antenna radiation resistance is for $L \gtrsim\!> \lambda$ given by[18,21]:

$$R_a = 60\ (1.4 + \ln(2L/\lambda))\ \Omega \qquad (4)$$

and typically is of the order of 100 Ω, so that $g_m \sim L/\lambda$.

In (4), resistive losses of the antenna due to the skin effect have been neglected. At low temperatures however, the electron mean free path l or scattering time τ_0 are much larger than at room temperature ($l \sim 5\mu m$, $\tau_0 \sim 5\cdot 10^{-12}$ s for the copper used). Therefore, even at far-infrared frequencies, the metal is already in the relaxation region ($\omega\tau_0 > 1$) and even in the anomalous skin effect region ($l > \delta$, δ skin depth), so that internal losses might become relevant. Contribution to the antenna impedance in these regions have been considered[22,23] and it can be shown[24] that they are in the order of a few Ohm at most for the relevant parameters. Internal losses are thus still negligible compared to R_a.

For the present experimental set up, estimates for the total cross section (as integrated over the total solid angle) are impossible to make, because the intensity distribution is unknown after the radiation has traversed the lightpipe and cone. It was noticed, that the detected signal is very sensitive to adjustments of the coupling of the laser beam to the entrance of the lightpipe (with only weak changes in transmitted power as observed by the monitoring detector). The direction of incidence is expected to affect also the angular intensity distribution at the location of the point contact. It should be noted that lightpipe wall reflections and especially reflection against the cone wall, which is able to change the propagation direction of the radiation, are important: for long antennas, the direction of the main lobe makes a small angle with the antenna (25° for $L/\lambda=4$). In addition, in the wavelength range considered, wavelengths ($\sim$ 0.5 mm) approach the dimension of the lightpipe ($\sim$4 mm) so that diffraction effects are very important. The dimension of the lightpipe however, is still too much oversized to make a waveguide analysis of the coupling problem tractable[25]. In principle, these coupling problems are easily circumvented by using a dewar with windows, giving direct optical access to the junction. Such an approach, which often is chosen when dealing with high-frequency Josephson effect studies[26], would merely go at the cost of a great deal of the simplicity of the present system.

For simplicity, we will write the effective total cross section to be used in (1) just as:

$$\sigma_a = (\lambda^2/4\pi)g_{eff} \tag{5}$$

g_{eff} is expected to be proportional to $g_m \cdot \Omega$ where Ω is the solid angle subtended by the antenna lobes. As $\Omega \sim \lambda/L$, g_{eff} is estimated to be independent of L/λ (or L and λ), supposed that the angular intensity distribution is independent of wavelength.

The spreading resistance is usually assumed to be Ohmic and given by the well known expression[27]:

$$R_s = \rho/2a \tag{6}$$

where a is the diameter of the contact and ρ the resistivity of the metal ($\rho \equiv 1/\sigma$).

The capacitance C is determined by the contact area and the thickness of the dielectric layer for MIM junctions or the (voltage dependent) thickness of the depletion layer for a Schottky diode[27].

The resistance R_b is determined by the contact barrier. Conduction across the barrier is due to thermal excitation in metal-semiconductor contacts and is assumed to be caused by direct quantummechanical tunneling for MIM diodes. The useful nonlinear properties are contained in R_b. To derive the nonlinear I(V) characteristic of MIM junctions, based on tunneling, together with a comparison of experimental high-frequency behaviour, has been the subject of a great deal of papers and is still controversial[13]. Interpretation of the high-frequency response has been based on tunnelmodels for planar tunneljunction geometries[4], direct laser enhanced tunneling due to the specific effects of laser heating and high laser-induced electric fields[28,11] and on special influence of the point contact geometry on the tunneling behaviour[29].

2. 3 ii. Nonlinearity of the spreading resistance.

The present work is concerned with rectification due to a nonlinearity of the spreading resistance only, where the barrier resistance may be completely absent or Ohmic. The nonlinearity of R_s has a simple physical origin[14]: the resistivity in a simple Drude expression is given by:

$$\rho = \frac{m^*}{ne^2} \cdot \frac{1}{\tau} \tag{7}$$

with m^* electron effective mass, n number density and e electron charge. The relaxation time τ is in general determined by impurity- and lattice scattering and can be written as:

$$1/\tau = 1/\tau_{imp} + 1/\tau_1 \tag{8}$$

with τ_{imp} impurity scattering time and τ_1 lattice scattering time. The lattice scattering time due to electron-phonon scattering depends on energy because of the energy dependence of the phonon density of states as well as of the interaction matrix element. The resistance therefore depends on voltage when the electrons are injected across the contact at an energy eV above equilibrium.

Investigations of this type of nonlinearities in point contacts between metals are used to study electron-phonon interaction and other relaxation processes in metals[14,30]. The nonlinearity in the I(V) characteristic is largest when the relative contribution of R_s to the total resistance is large. This situation is obtained in a pure constriction type contact, that is two bulk metals connected by a spherical orifice of radius a. For large a, the total resistance is formed by the spreading resistance only and the contact is Ohmic, as electron flow is diffusive. This is no longer true under conditions of ballistic electron transport, which arises when the electronic mean free path 1 becomes comparable to, or larger than, a: the electric field is confined to a region close to the orifice and electrons traverse the field region without scattering and are accelerated to an energy eV. For 1 >> a, the resistance is given by[31]:

$$R_{FE} = \frac{4}{3} \frac{1\rho}{\pi a^2} \tag{9}$$

For a narrow constriction $(a \ll 1)$, current is completely controlled by a field emission (FE) process and in this respect a constriction behaves similar to a tunneljunction. Note that now $l\rho$ is independent of the mean free path and thus R_{FE} is independent of voltage. For an arbitrary value of the Knudsen number $K=l/a$, the total resistance of a constriction contact R_c is given by an interpolation formula due to Wexler[30,32]:

$$\begin{aligned} R_c &= \Gamma(K)R_s + R_{FE} \\ &= \tfrac{4}{3}(1/\pi a^2\sigma)\left(1+\tfrac{3\pi}{8}\,\Gamma(K)\tfrac{a}{l}\right) \end{aligned} \tag{10}$$

Here, $\Gamma(K)$ is a slowly varying function of K, ranging from $\Gamma(K)=1$ at $K=0$ to $\Gamma(K)=0.694$ at $K=\infty$. Note that the nonlinearity of R_c is the nonlinearity of $\Gamma(K)R_s$ only (R_{FE} being constant); the field emission effect however, is essential to accelerate the electrons.

Using known expressions for $\tau_1=\tau_1$(energy) and combining (7),(8) and (10), the nonlinearity can be expressed as (at $T=0$)[14]:

$$\frac{dR_c}{dV} \cong \Gamma(K)\,\frac{dR_s}{dV} = \frac{3\pi^2}{4}\,R_{FE}\,\frac{ae}{\hbar v_F}\,\Gamma(K)\,\alpha^2F(eV) \tag{11}$$

V_1 is the voltage, $\hbar$ Planc constant/2π, v_F the Fermi velocity and where F is the phonon density of states and α^2 an effective interaction matrix element. The voltage derivative of the resistance thus directly reflects the Eliashberg function α^2F, which is strongly peaked at the energies of the Brillouin zone boundary phonons ($\sim$Debye energy).

A more rigorous and microscopic theory of the constriction resistance also exists[33] and the results are consistent with the simple description given above. The nonlinearities are found as corrections to the field emission current as a result of scattering processes, which scatter the electrons back through the hole.

Fig. 3 gives a typical "point contact spectrum" for a Cu-Cu point contact at $T=1.5$ K. The resistance is seen to increase strongly at higher bias voltages. The peak structure in the second derivative $d^2V/dI^2=R(dR/dV)$ displays the structure in α^2F: the large (broadened) peak at 17 meV corresponds to the two transverse phonon frequencies in Cu, while the second peak is in agreement with the longitudinal phonon frequencies. A "background" signal in d^2V/dI^2 is present in the experimental curves[14]. Note that the function F in (11) disappears above the Debye energy. The background responsible for the nonzero value of d^2V/dI^2 above the Debye energy has been explained by a nonequilibrium phonon distribution near the orifice[30]. Stimulated emission

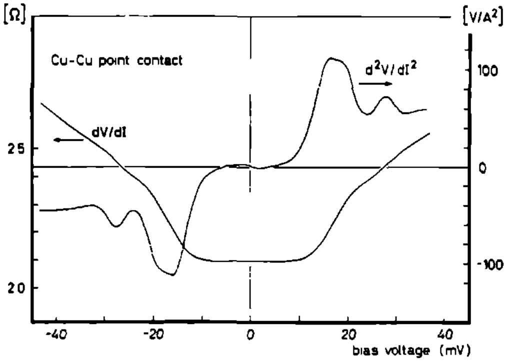
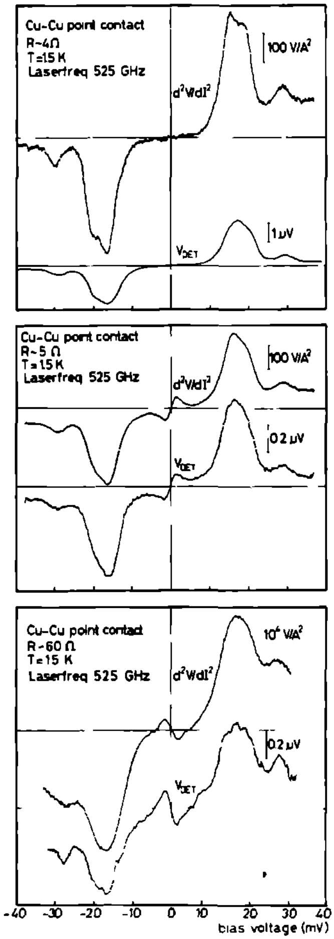
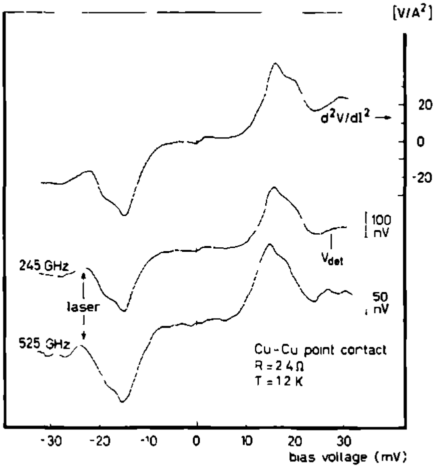


Fig. 3. Typical first and second derivative of the I(V) characteristic for a point contact between normal metals at low temperature.



(a)



(b)

Fig 4. (a): Second derivative (d^2V/dI^2) *and laser detected signal* (V_{det}) *curves as a function of bias voltage for three different contacts. (b): Second derivative* (d^2V/dI^2) *and laser detected signal* (V_{det}) *curves as a function of bias voltage at two laserfrequencies.*

and absorption of phonons leads to additional backscattering of electrons, yielding a continuing increase at high bias voltages. The magnitude of the background is found to vary from contact to contact[14].

It is important to note that the constriction resistance for pure contacts ($l \gg a$) is determined by R_{FE}. From (10):

$$R_{FE}/R_s = \frac{8}{3\pi} \cdot \frac{l}{a} \tag{12}$$

Values for l applying to the present contacts at low temperature may be about 5 μm, while it is generally accepted that typical constriction radii (as, in fact, deduced from (9)) for normal metal contacts[14] and for Josephson point contacts[34] may go down to 30-50 Å. Even at room temperature, in reasonably (but not exceptionally) pure metals, mean free paths may be at least as large as 30-50 Å. The resistance of a "shorted" contact may therefore under some conditions be much higher than expected from the spreading resistance only[14,34].

A useful measure for the nonlinearity of an I(V) characteristic is the logarithmic slope S:

$$S = \frac{1}{R}\frac{dR}{dV} = \frac{d^2V/dI^2}{(dV/dI)^2} = \frac{d^2I/dV^2}{dI/dV} \tag{13}$$

The nonlinearity for a pure constriction type contact S_c is then given by:

$$S_c = \frac{3\pi^2}{4} \frac{ae}{\hbar v_F} \Gamma(K)\, \alpha^2 F \cdot \frac{R_{FE}}{\Gamma(K) R_s + R_{FE}} \tag{14}$$

which holds only for $R_{FE} \gg R_s$, otherwise the electrons are not accelerated and (11) is unphysical.

Thus $S \propto a \propto R_c^{-\frac{1}{2}}$. Typical measured values for S at optimum bias voltage at low temperature are in the order of 10 V^{-1} for contact resistances in the order of 10 Ω (see Fig. 3).

The well known[14] results for the constriction resistance have been summarized here to some extent in order to decide what the equivalent high-frequency representation of a pure constriction would be. From the foregoing, it will be clear that the resistance R_{FE}, responsible for field emission, is equivalent to R_b in Fig. 2. The spreading resistance for a pure constriction should then, according to (10) be replaced by R_s':

$$R_s' = \Gamma(K) R_s \tag{15}$$

$R_s' \sim R_s$ because $\Gamma(K) \sim 1$.

When a barrier does exist at the interface between the two metals, then the total resistance R_T of the contact is just:

$$R_T = R_s + R_b \tag{16}$$

Supposing the barrier thickness to be small compared to l and a, the nonlinearity S_T is derived in precisely the same way as before, the barrier now being responsible for field emission, assuming R_b constant:

$$S_T = \frac{3\pi^2}{4} \frac{ae}{\hbar v_F} \alpha^2 F \cdot \frac{R_{FE}}{R_s + R_b} \tag{17}$$

(Here R_{FE} simply is the expression given by (9), but has no direct physical meaning). It is assumed that the voltage drop is completely across the barrier, which implies $R_b \underset{\sim}{>>} R_s$.

It is interesting to compare the nonlinearities due to the spreading resistance only, in case of a pure constriction type and of a tunneljunction, both contacts having the same diameter:

$$\frac{S_c}{S_T} = \frac{\Gamma(K)(R_s + R_b)}{\Gamma(K)R_s + R_{FE}} \simeq \frac{\Gamma(K)R_b}{R_{FE}} \tag{18}$$

where the last approximation holds because $R_{FE} >> R_s$, $R_b >> R_s$.

It is easy to show[35] that (apart from the possible averaging factor of order unity) $R_b/R_{FE} = 1/D$, where D is the tunneling probability (assumed constant for low bias voltage). It is more useful to compare constriction and tunnelcontacts, which both have the same resistance. From (14), (17) and (18) it follows $S_c(R)/S_T(R) = (a_c/a_T)(1/D)$, where the constriction and tunnelcontact diameters a_c and a_T are related by the constraint that both contacts have the same resistance: $(1/a_c^2)/(1/a_T^2 D) = 1$, so that $a_c/a_T = D^{-\frac{1}{2}}$. Then:

$$\frac{S_T(R)}{S_c(R)} = D^{\frac{1}{2}} \tag{19}$$

The same result is found from the microscopic theory also[30]. The nonlinearity therefore becomes small when the tunneling probability decreases, but depends only weakly on it (as $D^{\frac{1}{2}}$).

The very small constriction radii, which are sometimes assumed in point contact spectroscopy[14,30] ($\lesssim 50$ Å) for high Ohmic junctions ($\gtrsim 10\ \Omega$), should not be identified with the physical contact diameter. The resistance in such contacts is determined by metallic channels which pierce through a surrounding oxide layer whose (tunnel) resistance is large compared to the resistance of the channel. The actual contact was generally made and optimized by a combination of mechanical adjustment and deliberate shorting by sparking.

To represent the barrier resistance for a shorted barrier (constriction), the resistance R_b in Fig. 2 should be replaced by R_b', the parallel combination of R_b and R_{FE}. For most constrictions, $R_{FE} << R_b$ so that $R_b'=R_{FE}$. The capacitance is determined by the contact area, which may be much larger than the constriction area. A tunneljunction resistance might also actually be determined by an area much smaller than the contact area, if the oxide layer is accidentally thin over a limited area, e.g. at a protuberance of either whisker or post. It is not clear whether this is likely to happen, at contact pressures which may exceed the plastic deformation limit.

2. 3 iii. High-frequency analysis.

To proceed, the high-frequency induced voltage across the contact will be calculated. It will be assumed that R_b in Fig. 2 is constant, so that R_s is the only nonlinear element. $R_s<<R_b,R_a$, so that R_s can be very well considered to be current controlled. The induced dc voltage across R_s (i.e. the measured dc laser induced voltage) is then given by:

$$V_{dc} = \tfrac{1}{4}\,\frac{d^2V}{dI^2}\,i_0^{\,2} \tag{20}$$

where the current through R_s is assumed to be modulated as $I=I_0+i_0\cos\omega t$. I_0 is the dc bias current and ω the angular frequency of the radiation. It is assumed that i_0 is small enough that only terms up to second order in the series expansion of the current-voltage characteristic need to be retained. The second derivative d^2V/dI^2 in (20) is the same as which is measured in a dc experiment, because R_b is assumed to be a constant (i.e. $dR_s/dI=d(R_s+R_b)/dI$). Thus the laser detected signal as a function of bias voltage can be directly compared with the independently measured d^2V/dI^2 curve[15].

Equation (20) strictly is all we need to analyze the experimental results. The analysis of the network of Fig. 2 will be carried through somewhat further in order to gain some insight into the behaviour of this new type of high-frequency rectification. It is convenient to define S', the nonlinearity parameter determined by R_s only:

$$S' = \frac{1}{R_s}\frac{dR_s}{dV'} \tag{21}$$

By definition, the derivative must be taken with respect to the voltage dV' across R_s only. The ratio between dV'and dV, the total voltage differential across the junction can be estimated from Fig. 2:

$$\left|\frac{dV'}{dV}\right| = R_s\left\{\frac{1+(\omega CR_b)^2}{(R_s+R_b)^2 + (\omega CR_bR_s)^2}\right\}^{\frac{1}{2}} \tag{22}$$

Some caution however is required in interpreting (22). The nonlinearity of R_s essentially depends on the voltage across R_b, so that it is not evident how to handle when R_b is shorted out by C at high frequencies. It turns out that (22) does in fact produce a frequency roll-off, which would not have been obtained by neglecting C in (22). The exact functional frequency dependence however should not be taken too serious.

We then have:

$$S' \equiv \frac{1}{R_s}\frac{dR_s}{dV'} = \frac{1}{R_s}\frac{dR_s}{dV}\frac{dV}{dV'}, = \frac{1}{R_s}\frac{dR_s}{dV}\ \frac{1}{R_s}\left\{\frac{(R_s+R_b)^2+(\omega CR_bR_s)^2}{1+(\omega CR_b)^2}\right\}^{\frac{1}{2}} \tag{23}$$

From the definition of the experimentally measured, total nonlinearity parameter, S,of the junction, eq. (14) or (18), we have:

$$\frac{1}{R_s}\frac{dR_s}{dV} = \frac{R_s+R_b}{R_s}\ \frac{1}{R_s+R_b}\ \frac{d(R_s+R_b)}{dV} = \frac{R_s+R_b}{R_s}\ S \tag{24}$$

so that the relation between S' and S becomes:

$$S' = \frac{R_s+R_b}{R_s^{\ 2}}\left\{\frac{(R_s+R_b)^2 + (\omega CR_bR_s)^2}{1+(\omega CR_b)^2}\right\}^{\frac{1}{2}}\cdot S \tag{25}$$

The power transfer function F, defined as the power dissipated in the nonlinear element, P_s ($\sim v_s^{\ 2}/2R_s$, with v_s high-frequency voltage amplitude across R_s) divided by the power P_a which can be delivered by the generator ($\equiv v_a^{\ 2}/8R_a$) follows from the network of Fig. 2 as:

$$F = \frac{4R_aR_s(1+(\omega CR_b)^2)}{(R_a+R_b+R_s)^2 + (\omega CR_b(R_a+R_s))^2} \tag{26}$$

At high frequencies, F is simply given by $4R_aR_s/(R_a+R_s)^2$ as expected, independent of frequency. The nonlinearity parameter gives the current sensitivity of a nonlinear element for small signals ($\frac{1}{2}S'$). The voltage responsivity defined as the dc induced voltage per unit power absorbed is therefore given by $\frac{1}{2}S'R_s$. For the (open circuit) voltage we find, combining (25) and (26):

$$V_{dc} = \tfrac{1}{2}S(R_s+R_b)\cdot\frac{4R_a\{(R_s+R_b)^2+(\omega C R_b R_s)^2\}^{\frac{1}{2}}}{(R_a+R_b+R_s)^2+(\omega C R_b(R_a+R_s))^2}\cdot P_a \qquad (27)$$

At high frequencies, this gives a roll-off as $1/\omega$, but as remarked, the exact functional dependence on ω may be incorrect.

Typical estimated values for C, based on estimated contact area and thickness of the insulating layer, are in the order of 10^{-15}F or less. At the frequencies of interest, it seems reasonable to ignore the capacitance. For a tunneljunction, the product R_bC is independent on contact area because $C \propto a^2$ and $R_b \propto 1/a^2$. For a pure constriction contact, R_b and C are more independent because the constriction radius ($R_{FE} \propto 1/a^2$ also) will mostly be much smaller than the contact area. Negelecting C, (27) becomes:

$$V_{dc} = \tfrac{1}{2}\ S\ (R_s+R_b)\cdot\{\frac{4R_a(R_s+R_b)}{(R_a+(R_s+R_b))^2}\}\cdot P_a \qquad (28)$$

which is the expected result.

Eq. (28) could have been obtained immediately by treating the junction as a single, nonlinear resistive element, with a resistance value (R_s+R_b) and nonlinearity parameter S. The factor in curled parentheses can now be viewed as the power transfer function. Impedance matching corresponds to $R_a=R_s+R_b$. Because $R_b >> R_s$, R_s can be neglected in (28). The spreading resistance is usually neglected for MIM junctions, although a priori. Eq. (28) in fact is exactly the same expression which would have been obtained if the barrier resistance was nonlinear, with nonlinearity parameter S.

This result might not have been anticipated from the equivalent circuit of Fig. 2. Because the value of R_s is negligibly small compared to R_a and R_b, negligible power is dissipated in the nonlinear element. The reason for the nevertheless important effect of R_s is due to its intrinsic very strong nonlinearity S' and the strong dependence of S' on R_s ($S' \propto 1/R_s^2$, eq. (25)), which precisely cancels both the effects of power transfer mismatch (eq. (26)) as well as the reduced rectified voltage (via $V_{dc} = \frac{1}{2}S'R$). It must be emphasized that (28) holds, both for pure constriction type contacts as well as for barrier type of contacts (e.g. tunneling), with the appropriate values for S in both cases.

The rather trivial result (28) shows the internal consistency of treating the nonlinear spreading resistance in a standard equivalent circuit representation of Fig. 2 and the, somewhat artificial and experimentally unattainable, use of the nonlinearity parameter S' (eq. (23)). Still it should be kept in mind that it is by no means evident that the highly complicated process, responsible for the nonlinearity, especially in case of a constriction resistance, can adequately be described in terms of a distribution of (nonlinear) lumped impedances. Such model is expected to be clearly inadequate to predict the high-frequency roll-off due to external circuit parameters. The validity of the use of the equivalent circuit representation and the resulting formulas equivalent to (28) (but including the frequency dependence) have been extensively tested in the literature for MIM diodes[5,17-19].

2.4. Experimental results.

2.4 i. Typical data for constriction type contacts at low temperature.

A typical set of data, obtained for three different contacts at liquid helium temperatures, is shown in Fig. 4(a). The directly measured second derivative d^2V/dI^2 and the laser detected signal are given as a function of bias voltage. As can be seen, the detected signal is in excellent agreement with the d^2V/dI^2 characteristic. The detection mechanism is clearly due to rectification[15] and in agreement with the simple expression (20). The main nonlinearity is caused by the nonlinear spreading resistance as discussed, but the three contacts of Fig. 4 differ by their behaviour near zero bias voltage.

A so called "zero bias anomaly" is frequently observed in this type of contacts[14] and appears as either a resistance minimum or maximum (see Fig. 4). Its physical origin is not very well established. For constriction type contacts, a mechanism based on the Kondo effect explained the resistance maxima observed in point contacts between noble metals, which were doped with magnetic impurities[36]. The nonlinearity in this case results also from a bulk scattering process and thus essentially is caused by the spreading resistance. Zero bias anomalies are well known from tunneling spectroscopy using planar tunneljunctions also[37], but are then attributed to R_b. For a general point contact such as those of Fig. 4(a), it is not known whether the nonlinearity should be attributed to R_s or R_b (c.q. R_{FE} for a constriction). In any case, rectification at the zero bias anomalies takes place just as well as for the main nonlinearity. In fact, this result is of interest per se, because it suggests that the mechanism responsible for the zero bias anomaly is of microscopic nature, being inherently fast and not a spurious effect of thermal or other nature. Such spurious effects are not likely to give precisely the same bias voltage dependence for the measured d^2V/dI^2-curve and the laser induced signal, when normalized with respect to the phonon scattering induced peak in both cases.

As expected for rectification, the results are independent of the frequency (see Fig. 4(b)). Slight differences in the two curves taken at 245 and 525 GHz laserfrequency should be attributed to accidental laserpower instabilities.

Two examples of rectification observed at extreme (constriction) types of point contacts are shown in Fig. 5. One contact has an exceptionally large zero bias anomaly, whose nonlinearity exceeds those of the phonon peak. In the laser detected signal, the zero bias anomaly appears somewhat too small, but this may be attributed to a broadening effect. The high-frequency voltage amplitude for this case can be calculated to be about 2 mV, so that the narrow top is somewhat smeared out. The junction employed for the right hand side of Fig. 5 has an exceptionally large "background", that is the second derivative d^2V/dI^2 is large even for voltages in excess of the phonon energies. Although there is no obvious dependence on the resistance, a high background is usually associated with low resistances[30]. The result obtained with the high background contact has also significance in that it helps to rule out detection by spurious effcts, such as heating, because it is known that the temperature dependence of the I(V) characteristic at the phonon energies differs from that at higher voltages[30].

As expected from the calculated responsivity (eq. (28)), most sensitive junctions are obtained for high resistances. These therefore enable measurements to be made as a function of power with the available laserpower, even with the rather poor coupling scheme employed. Typical results are given in Fig. 6, where the output signal (and so also roughly the dissipated laser power) changes over two decades. The laser beam was attenuated by inserting lossy dielectrics (carton) in the beam, but the attenuation was not calibrated. For increasing power, the high-frequency voltage swing across the contact becomes larger than the width of the structure in the I(V) characteristic and so structure is smeared out.

Assuming that the high-frequency voltage amplitude v_1 can still be estimated from $V_{dc}=\frac{1}{4}Sv_1^2$, with the, independently measured, value for S (5 V^{-1} for this contact at the maximum), v_1 is found to be 1, 3 and 8 mV in the three

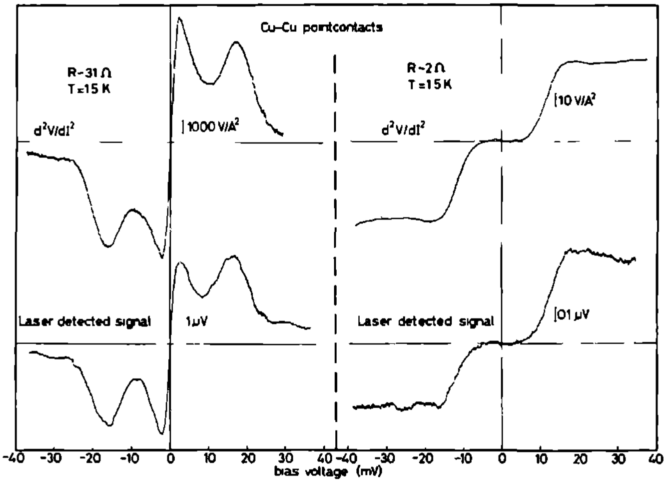


Fig. 5. Second derivative (d^2V/dI^2) and laser detected signal (V_{det}) curves as a function of bias voltage for a contact with an exceptionally high zero-bias anomaly (left) and a contact with an exceptionally high "background" (right).

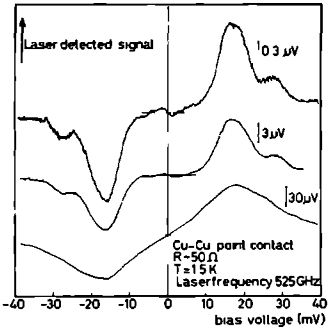


Fig. 6. Laser detected signal curves for a sensitive (high-Ohmic) point contact for three different laser powers, showing the effect of power broadening.

cases. Standard rules give a "modulation broadening resolution"[30] of $1.7v_1$, which should be compared to the intrinsic width of the structure (∿10 mV). The observed width of the structure of 9.5, 11 and 22 mV is thus very well consistent with the independent estimates for v_1. (A direct comparison with low-frequency overmodulation was not practical in the present set-up and has been omitted).

For junctions which accidentally have a zero bias anomaly which is reverse in sign with respect to the main peak, overmodulation will lead to a polarity reversal of the detected signal when the power level increases, for a limited range of dc bias voltage[15]. Actually, such a polarity reversal also is present in the data of Fig. 6, for dc bias voltage near zero. In most of the other measurements described in this work, the laserpower was kept sufficiently low, such that the shape of the output signal against bias voltage curve did not depend on power.

2. 4 ii. Data as a function of temperature.

To enable a better comparison of the presently described rectification mechanism with more conventional MIM diodes (or perhaps also Schottky diodes) which operate at room temperature, measurements have been made at several temperatures (Fig. 7). Both the measured d^2V/dI^2 characteristic and the laser induced signal plot for the same contact are shown at liquid helium, liquid nitrogen and room temperature and at ∿200 K. For temperatures above those of liquid helium, the contact was in ambient gas atmosphere (He) and not in a liquid. Because there is a necessary time delay between the temperature points, the contact was readjusted at each temperature (but the same whisker and post were used). Also, the coupling lightpipe between laser and dewar was removed between different temperature measurements, so that power and especially power coupling to the whisker may be different in each case. The obtained sensitivities can therefore not directly be compared. The temperature dependence of the I(V) characteristic itself is well documented in the literature[38] and the d^2V/dI^2 curves shown in Fig. 7 are well representative for it, although taken on different contacts. The distinghuished structure in d^2V/dI^2 at low temperatures gradually broadens and finally disappears with increasing temperature. It was shown[38] that the temperature limited resolution for the phonon peaks in constriction contacts amounts 5.4 kT (k Boltzmann constant), equal to the temperature broadening in electron tunneling spectroscopy. (Note that kT∿25 meV at room temperature).

The magnitude of the "background", which is constant at high voltage, does not depend on temperature. The contacts of Fig. 7 are all low-Ohmic ($\sim 7\ \Omega$) and all believed to be of the pure constriction type. It can be seen that the nonlinearity parameter S ($\equiv (d^2V/dI^2)/(dV/dI)^2$) is approximately constant ($\sim 2\ V^{-1}$) for all contacts shown, at the maximum applied bias voltage. The gradual broadening, which tends to a constant level for d^2V/dI^2 at high bias voltage results in a nearly straight line at room temperature. The saturation to the constant background level at room temperature is expected to occur outside the voltage range displayed in Fig. 7. For rather low Ohmic junctions of the constriction type, current densities at the constriction will exceed values as high as $10^{10}\ A/cm^2$ in the tens of mV bias range. It was found that junctions of this type tend to become very noisy, unstable and apt to disrupture of the contact when bias voltages approach 100 mV and generally cannot be biased at voltages significantly in excess of 100 mV. These effects are believed to be related to the high current densities, which also lead to excessive heating.

The laser detected signal curves, shown also in Fig. 7 display precisely the same evolution with temperature as the second derivative, which suggests that detection by rectification continues to be the dominant detection mechanism up to room temperature and the nonlinearity can still be ascribed to the spreading resistance. At liquid helium temperatures, the polarity of the rectified voltage could be verified, but this could not be done directly at the higher temperatures. The phase setting of the lock-in amplifier however, was adjusted for the low temperature data.

A characteristic difference between the presently discussed I(V) characteristic (where the nonlinearity is essentially due to the spreading resistance) and a tunneling characteristic (as assumed for MIM diodes) is the curvature of the I(V) characteristic, which is opposite in both cases. For the first type, the differential resistance increases with applied bias voltages (e.g. see Fig. 3), whereas tunneling models predict a decrease. Numerous published calculated I(V) curves, based on tunneling models show this[39,4,13]. Published experimental I(V) characteristics for point contacts[4,5] or small area evaporated tunneljunctions[40,41] do in general show a curvature consistent with tunneling models. However, some of the published I(V) characteristics for low-Ohmic junctions do in fact show the opposite curvature[42] (see also curve I in Fig. 3 of ref. 4). Many experimental studies of rectification[4,43,44] as a function of bias voltage are performed at high power (i.e. high modulation amplitude), so that detected signals can be of either

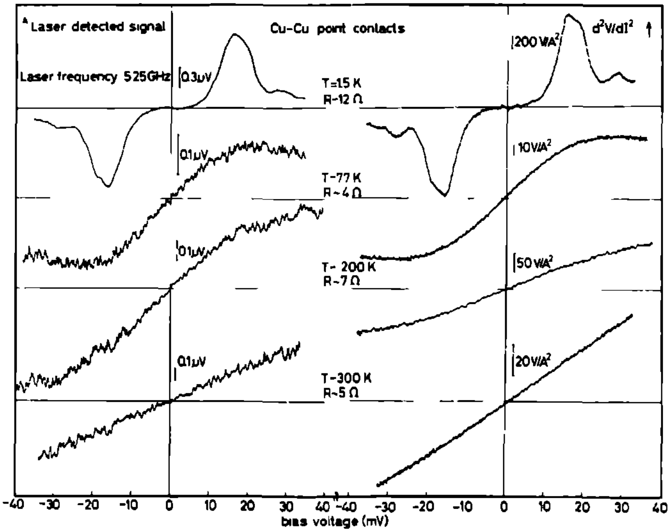


Fig. 7. Laser detected signal curves (left) and second derivative (d^2V/dI^2) curves (right) as a function of bias voltage at different temperatures. For each temperature, the contact was newly adjusted.

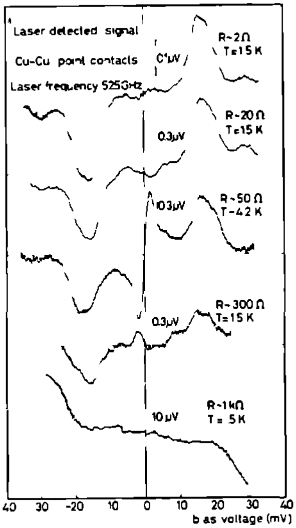


Fig. 8. Laser detected signal curves as a function of bias voltage for contacts with varying resistances.

polarity (depending on parameters of the tunnelcharacteristic and magnitude of modulation amplitude), irrespective of the actual sign of the curvature ($\sim d^2V/dI^2$) at the dc bias point. Calculation of large signal rectification and mixing, including the associated polarity reversals, based on parametrized tunneling I(V) characteristics, have been extensively published[4,39,43-45]. For low power levels however, the polarity of the rectified signal for a tunneldiode will be opposed to that for the presently discussed nonlinearity[46]. A reverse polarity of the detected signal for low-Ohmic junctions as compared to high-Ohmic junctions has been reported indeed[47,4] at a wavelength of 1.15 μm. Although such results are in qualitative agreement with tunneling models, a rigorous verification is difficult to obtain, because the I(V) characteristic strongly depends upon unknown parameters, while the induced optical voltage is difficult to estimate independently. From the detection results of Fig. 7, which are consistent with the I(V) characteristic, it is concluded, that the spreading resistance will be responsible for rectification also at room temperature, in cases where the tunneling nonlinearity is negligibly small. The dc I(V) characteristic is clearly inconsistent with tunneling theory, because of the sign of its curvature. Irrespective of the influence of R_s in practical MIM devices, its effect must evidently be taken into account in systematic investigations of the diode itself, which might include less sensitive[4,42] low-Ohmic contacts.

2. 4 iii. Data as a function of contact resistance.

A series of detection results, obtained for junctions of differing resistance values are given in Fig. 8. Each resistance is independently adjusted and the data form a representative, although selected, view of commonly observed behaviour. In the present set-up, high-Ohmic contacts ($\gg$ 100 Ω) are somewhat difficult to obtain and are generally unstable and very sensitive to ambient electrical disturbances. Direct measurements of the I(V) characteristics or second derivatives, in addition to the detection measurements have therefore not been made.

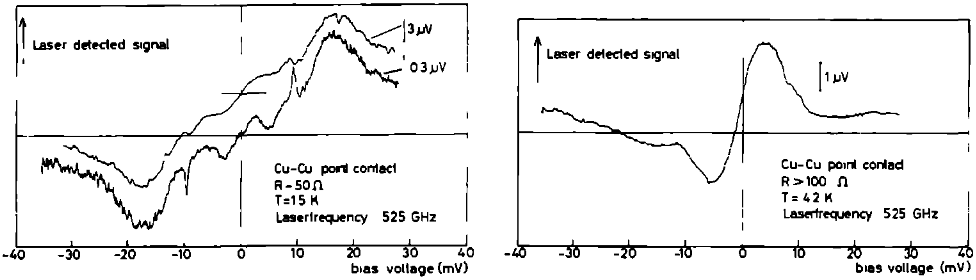
The low-Ohmic contacts ($\lesssim$ 50 Ω) invariably have the phonon scattering induced peak as the dominant nonlinearity. For higher Ohmic junctions, the rectification behaviour (and thus shape of the I(V) characteristic) varies from contact to contact in a rather unpredictable way. The phonon peaks still are however frequently the dominant nonlinearity in contacts of several hundred Ohm (see 300 Ω result, Fig. 8).

A very interesting behaviour is displayed by the high-Ohmic contact in the lower trace of Fig. 8. The polarity of the signal has now reversed with respect to the upper traces. It is very suggestive to decompose the resulting lineshape in the normal, phononscattering induced signal, superimposed on a straight line with negative slope. Such a decomposition would be consistent with a model in which both the spreading resistance is nonlinear (in the way discussed in this work), as well as the barrier resistance. The negative slope of the barrier resistance nonlinearity is consistent with tunneling models and a linear variation with bias voltage is expected at small bias voltage.

It is very hard to imagine that high-Ohmic contacts can still be of the constriction type. According to the expression for R_{FE}, unphysically small contact diameters would be required: $a \sim 4$ Å for $R_{FE} \sim 1$ kΩ.In addition, such radius also is of the order of the De Broglie wavelength, so that the theory no longer holds. When scaled to typical values at low resistance according to (14), S_c values in the order of $\lesssim 1\ V^{-1}$ would be expected for the (academic) case of a 1 kΩ constriction contact ($S_c \sim R_{FE}^{-\frac{1}{2}}$) at the maximum. Measured S values unfortunately are not available for the high-Ohmic junction of Fig. 8. A lower limit however can be given by noting from the 1 kΩ result that the curve is sharp on a 10 mV scale, so that the induced high frequency sweep v_1 must be limited to $\lesssim$ 10 mV. From $V_{dc}=\frac{1}{4}Sv_1^2$, it is then found $S \gtrsim 0.5\ V^{-1}$ (In fact, the estimate $v_1 \lesssim$ 10 mV is consistent with the previous estimates for high-Ohmic junctions).

The nonlinearity S_T caused by the spreading resistance in case of a tunneljunction will be of order $D^{\frac{1}{2}}S_c$ (eq. (19)). Here, D is the electron tunneling probability and S_c the nonlinearity of the (academic) constriction giving rise to the same resistance as the tunneljunction. For $R \sim 1$kΩ , $S_c \sim 1\ V^{-1}$, it must be required $D \gtrsim 0.25$ to arrive at the estimated $S_T \gtrsim 0.5\ V^{-1}$ value. These rough estimates serve only to show that the assumption of a non-linear spreading resistance for a tunneljunction does not lead to unphysical results, but more rigorous arguments are impossible to give. The (physically relevant) contact radius a is unknown, which holds even more for the tunneling probability, which depends on oxide thickness, for which estimates cannot be given. These problems are common to all tunnel-type point contact junctions and illustrate the difficulty to explain the widely varying types of published I(V) characteristics.

It should be noted that estimated values of S for high-Ohmic junctions are much smaller than at low resistances, so that high-Ohmic contacts are



(a) *(b)*

Fig. 9. Laser detected signal as a function of bias voltage for two different high resistance contacts showing anomalous behaviour.
(a): the sharp peaks near 10 mV bias are unexplained; at high power level they are smeared out.
(b): contact with high sensitivity in a low bias voltage region, with finite signal at zero bias.

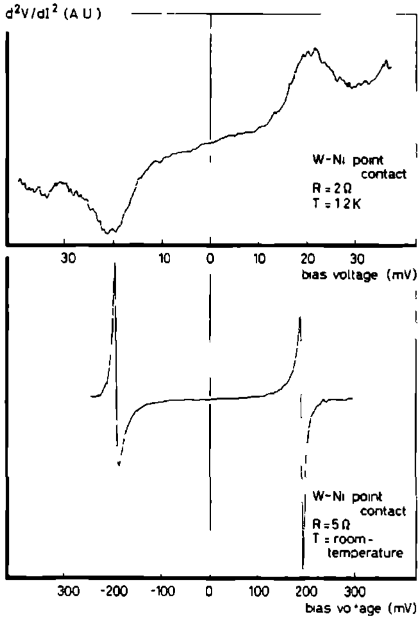


Fig. 10. Second derivative (d^2V/dI^2) curves of two W-Ni point contacts. The upper curve is taken at ∿1.2 K and displays the usual structure of a constriction type point contact. The lower curve is taken at room temperature and exhibits very strong nonlinearities near 200 mV bias voltage.

generally rejected for point contact spectroscopy[30]. In fact, low S-values in some cases might be more easily measured (though in arbitrary units) in high-frequency detection experiments than by low-frequency modulation and phase-sensitive detection techniques, as long as rectification dominates thermal (or other) effects. This is due to the fact that detection essentially depends on deviations from nonlinearity, while for a low-frequency modulation technique the nonlinearities have to be resolved above a linear background (e.g. by the bridge circuit).

2. 4 iv. Data for other types of contacts.

Fig. 9 serves to illustrate the variety of signal shapes, which may accidentally occur in high resistance contacts. Because anomalous behaviour is only observed for high resistances, it is likely that these effects should be attributed to an interface effect (∿oxide- or impurity layer) and therefore to the barrier resistance rather than to the spreading resistance.

In Fig. 9(a), in addition to the spreading resistance nonlinearity, a narrow peak is superimposed at a bias voltage ± 10 mV. It might be the excitation of some vibrational impurity mode in the oxide (or metal surface) layer[37], although the energy would be somewhat small. The peak is smeared out at high modulation voltages (upper curve of Fig. 9(a)). A contact of 50 Ω is somewhat in the transition region, where both metallic constriction and oxide tunneling could be reasonable, so that the barrier effects may occur in parallel to the constriction.

In the contact of Fig. 9(b), the detection signal is not antisymmetric with respect to bias voltage, has a relatively large signal at zero bias and no effects due to the spreading resistance can be identified. The usual tunneling theories do neither explain such behaviour, because then the signal should be linear, at least for such low bias voltages and power levels. (Actually, the sign of the polarity was not verified in Fig. 9(b) and has been set arbitrary). Although it was found experimentally that tunneljunction characteristics significantly depend upon temperature[41], this is not evident from theory (e.g. ref. 39), as long as kT is small compared to the barrier height. For reasonable barrier heights (∿eV), this is the case at room temperature as well as at liquid helium temperature. Presumably, the structure in Fig. 9(b) is related to the (equally unexplained) zero bias anomalies discussed before.

2. 5. Discussion.

The subject of the present work is to investigate the high-frequency operation of a special type of point contact (notably at low temperature), where the nonlinearity is caused by an electron-phonon scattering process and must be ascribed to the spreading resistance instead of the barrier resistance. The physical origin of this type of nonlinearity is very well understood and has a rigorous theoretical basis. Because it is essentially due to a bulk process, it is a very pertinent and reproducible effect for not too high resistance junctions. Both these merits are in clear contrast with the nonlinearity of the barrier resistance due to elctron tunneling, which varies strongly from contact to contact and can be explained only by, sometimes rather arbitrary, adjustment of the tunneling-model parameters.

The spreading resistance nonlinearity therefore is useful for model investigations after the high-frequency behaviour of point contact devices and investigations of it may thus add support to the widely given confidence to other types of point contact devices, notably the so called MIM diode.

It is however useful to set out the potential merits or drawbacks for exploiting the mechanism described in this work by itself for some practical applications. For this purpose, it is just a matter of operating the standard MIM device in a mode which enhances the spreading resistance nonlinearity, while suppressing the tunnelresistance nonlinearity. This will generally imply application of higher pressure (low resistance) and a selection of electrode materials, while even operation at low temperature might be considered.

The potential relevance of the spreading resistance for arbitrary point contacts is shown by Fig. 10, curve a, where the low temperature d^2V/dI^2 curve of a W-Ni contact is displayed. The combination W-Ni is frequently used for practical high-frequency MIM diodes; apart from the superior mechanical properties of W, the difference in work functions of the metals W and Ni is large, which favours a large tunneling nonlinearity according to tunneling theory. The result shown in Fig. 10, curve a, clearly exhibits the (presumably superimposed[14]) well known point contact spectrum of W[48] and Ni[49]. On the other hand, the metal Ni appears to be not very representative for investigating low-Ohmic point contact behaviour, due to its magnetic properties. Structure in the point contact I(V) characteristic due to magnetic surface subbands[50] or heating, when the junction was biased at voltages which heated the contact up to near its ferromagnetic transition temperature[49] have been observed at liquid helium temperatures. The latter effect, which leads to large nonlinearities, also easily occurs at room temperature (see Fig.10, curve b). Some care therefore is required when Ni is used as one of the electrodes.

The most important figure of merit for a high-frequency diode obviously is the high-frequency cut-off, where extrinsic (circuit parameters) and intrinsic cut-offs must be distinghuished. The principal circuit limitation is often given by the time constant R_sC (especially in connection with Schottky diodes). For $\omega \gtrsim 1/R_sC$, the voltage drop over R_s (considered as a parasitic impedance) becomes larger than over R_b (considered the nonlinear element) and thus parasitic losses dominate. At high frequencies (far infrared or higher), the embedding network (simply represented by R_a in Fig. 2) becomes simple (open antenna configuration) and, moreover, uncontrolled. Thus power impedance mismatch between R_a and the junction will become prohibitive, because C cannot be tuned out. The time constant then is RC with R either R_a or R_b, whichever is lower[17,18]. For the present mechanism, it is clearly without sense to consider R_s as parasitic, while the interesting frequency range is "high". Moreover, as discussed before, the principle of the operation mechanism may be affected when R_b is shorted by C. The limiting circuit time constant seems to be thus R_bC in all cases. Thus, circuit time constant is completely comparable to that of MIM diodes (or even better for small R_b) and presumably superior to Schottky diodes where the parasitic losses even become large at $\nu \gtrsim 1$ THz. MIM diodes may be operated above cut-off (with $1/\omega$ roll-off[18] or faster[17]) , but the ω-dependence of the roll-off is not well established for the present mechanism (eq. (27) suggests $1/\omega$). With some accepted value of $C \sim 10^{-15}$ F and a value $R_b \sim 20\ \Omega$, circuit cut-off will be $\nu_c \sim 10^{13}$ Hz.

From the physical mechanism behind the nonlinearity, it will be clear that there is an intrinsic frequency cut-off also, which can be roughly estimated by $\omega \sim 1/\tau$, where τ is some average electron-phonon relaxation time constant, determined by the phonons of maximum energy. Typical values for τ may be estimated from[30] :

$$\frac{1}{\tau} \sim \frac{2\pi}{3}\, \lambda_{ep}\, \frac{E_D}{\hbar} \qquad (29)$$

where λ_{ep} is the electron-phonon mass renormalization constant and E_D the Debye energy. This leads to values in the order of 10^{-13} s. Values for the stimulated emission (giving rise to the background signal) are difficult to estimate, but are at least as short. Thus both extrinsic and intrinsic, the frequency cut-off for the presently discussed nonlinearity will be in the order of 10^{13} s[51]. This is considerably smaller than the intrinsic time constant of MIM diodes, where it is determined by the time of flight of an electron through the oxide layer.

In the following, we will assume operation below cut-off and ignore C in the equivalent circuit. It was shown, that the analysis becomes trivial then and the junction can be considered as a single nonlinear element (its resistance given by $R=R_s+R_b$), however principally without parasitics. For a pure constriction, $R_s \ll R_b \ll R_a$, and theoretically $S \propto R_b^{-\frac{1}{2}}$, so that the voltage sensitivity for video detection is proportional to $R^{3/2}$ (see (28)). This relation was not verified experimentally, because of the scatter in the S-values and coupling efficiency, but sensitivity was observed to increase considerably with resistance (presumably faster than linear). Best results therefore are expected with higher resistances, but of the pure constriction type. To evaluate performance quality of spreading resistance nonlinearity effects, pure constrictions need to be considered only, because tunnelling type contacts (with linear barrier !) will certainly be of less quality. Reasonable working values will be $R \sim 20\ \Omega$ and $S \sim 10\ V^{-1}$ at low temperature. For higher resistances, it is not clear that they will be of the pure constriction type and thus S is unreliable. Note that the impedance mismatch factor $F=4R_aR/(R_a+R)^2$ is as high as 40 % for such junction with a typical value $R_a=150\ \Omega$[17]. Impedance mismatch therefore is not a serious factor and is clearly comparable or even superior to high-Ohmic MIM junctions.

A useful measure for the sensitivity of a video detector is the noise equivalent power in W/ √Hz. Theoretically, noise will be limited by shot noise because 1/f noise can be overcome by choosing high enough modulation frequencies, while thermal (Johnson) noise is negligible compared to shot noise when $eV > 2kT$, which clearly is fulfilled at low temperatures. With the shot noise current given by $\langle i^2 \rangle = 2eI\Delta f$[27], where I is the dc bias current and Δf the bandwidth, the (voltage) NEP can be expressed as:

$$\mathrm{NEP} = \frac{\sqrt{2eI}}{\frac{1}{2}SF}\ \mathrm{W}/\sqrt{\mathrm{Hz}} \tag{30}$$

Inserting the values $I \sim 17$ mV/20Ω, $S=10\ V^{-1}$, $F=0.4$, NEP $\sim 10^{-11}$ W/√Hz (and a sensitivity of ∼40 V/W). A representative measured value for the sensitivity, when referring to the laser output power (∼1 mW estimated) was only $\sim 5\cdot 10^{-2}$ V/W. This means that only ∼0.1 % of the power is coupled at the antenna. Taking the intensity near the diode as $I=P_{\text{laser}}/\pi r^2$, with r the lower radius of the optical cone, we obtain for the antenna cross section in the natural units $\lambda^2/4\pi$ (see eq. 5), the value $4\pi\sigma_a/\lambda^2 \simeq 0.4$, which should be compared to 1. This indicates that the antenna coupling loss at this wavelength is primarily determined by lack of focussing, rather than by the directional properties of the antenna, as expected (see discussion around eq. (6)).

It should be noted that the parameters used for the estimate may be in fact somewhat conservative. Several of the contacts used in the figures of the preceding section have larger values for the product SF and lower current than used in (30).

Some care is required in comparing the quoted NEP-value with similar estimates for conventional room temperature MIM diodes. In practical cases, they are usually operated without external bias. Literature values for S at zero bias scatter by several orders of magnitude and from systematic investigations of the I(V) characteristic, S(V=0) appears to be generally a negligible fraction of the value at some optimum finite bias[4,5]. In fact, a large value of S at zero bias voltage is somewhat difficult to explain theoretically, especially for contacts between identical metals (a non-antisymmetric I(V) characteristic is required for nonzero S(V=0)). In recent work[17], a typical theoretical estimate for an unbiased room temperature MIM junction, using a measured S-value was given as NEP$\sim 10^{-10}$ W/$\sqrt{\text{Hz}}$. It was assumed that Johnson noise was dominant for the unbiased diode. Using however data from other work at finite bias[5] (the noise considered to be shot noise limited), a value $\lesssim 10^{-11}$ W/$\sqrt{\text{Hz}}$ may be inferred. Operating the above quoted unbiased diode[17] at liquid helium temperature, assuming that the diode parameters remain unchanged, would result also in a NEP$\lesssim 10^{-11}$W/$\sqrt{\text{Hz}}$. Thus it appears that sensitivity and noise properties of the low temperature constriction point contact are certainly comparable or perhaps superior to the best values which may be obtained for MIM diodes, even if they were operated at low temperature, with or without bias.

It should be added that the estimated performance for the constriction contact is obtained for a low resistance contact (∼20Ω), whereas for the typical tunnelcontacts the resistance is high (∼1kΩ). A low resistance contact has the indisputable advantage above a high resistance contact of being considerably more stable and reliable. In addition, the spreading resistance is very reproducible. Operation at liquid helium temperature is however a serious drawback. At room temperature the NEP for a constriction contact is expected to be reduced by at least an order of magnitude. The S-value will be lower (see Fig. 7), (roughly a factor of 5), and noise somewhat higher because bias current will be somewhat higher, while thermal noise may now be comparable to shot noise (factor 2). An interesting possibility exists in operation at liquid nitrogen temperature. For Schottky diodes (room temperature), measured NEP values for video detection in the order of 10^{-8} W/$\sqrt{\mathrm{Hz}}$[52] at FIR frequencies have been reported. After correction for antenna coupling losses, it is clear that these values will not be superior to the low temperature constriction diode.

A point contact device is however generally not interesting for video detection but only in a heterodyne mode of operation. To evaluate the mixer performance of a nonlinear element is a highly complex problem[27] and will not be attempted.

The comparison of video NEP for different detectors will give a first indication on their relative performance in a heterodyne application. From this number only, together with its relative stability, the low temperature constriction diode is competetive with and perhaps superior to, the MIM diode or Schottky diode. Because the primarily important frequency range of operation of MIM diodes is at near infrared or higher, the constriction diode will not be competetive with MIM diodes because of its expected low frequency cut-off. Photolithographically fabricated Schottky barrier diodes have presently demonstrated performance in the entire far-infrared range[52,53] and will perhaps operate at frequencies as high as 30 THz[54]. Impedance mismatch losses degrade their performance at high frequencies (>>1 THz)[55].

A very rough indication of the amount of local oscillator (LO) power required for minimum conversion loss for the constriction diode mixer will be given by the power level which drives the junction approximately Ohmic. The LO voltage sweep should then be in the order of the width over which the nonlinearity extends, which is some tens of mV. This requires a LO power in the order of 0.1 mW for the typical 20 Ω diode and compares with typical LO powers of Schottky diodes (mW range). Because at low temperature the nonlinearity is very sharp, higher order derivatives of the I(V) characteristic for the constriction are appreciable and the device might be capable of satisfactory higher order harmonic mixing which is often a useful property. Harmonic mixing will certainly strongly deteriorate at higher temperatures. Thus it seems that the constriction type diode is perhaps capable of competing with Schottky diodes at very high frequencies (but below $\sim 10^{13}$ Hz).

When working at liquid helium temperatures, a comparison with superconductive devices, of which the point contact Josephson junction [56] has the highest demonstrated frequency capabilities, must be made also. Heterodyne experiments and harmonic mixing with Josephson junctions are very well possible over the far-infrared region, but their intrinsic frequency limitation will increase the conversion losses at frequencies considerable in excess of 1 THz.

2. 6. Conclusion.

Radiation detection by metal-metal point contacts at low temperature has been investigated at far-infrared frequencies.All results are understood in terms of a rectification process of high-frequency currents, which are generated across the contact. The nonlinearity of interest in the present work is caused by the spreading resistance and not by the barrier. This effect is very strong at low temperatures for low-Ohmic contacts and dominates the detection mechanism. In principle, the mechanism also is active at room temperature and/or in high-Ohmic contacts, although much smaller. For very high-Ohmic junctions, another type of nonlinearity, presumably due to tunneling, although poorly understood, appears to be dominant.

Although the figures of merit of the diode for considering to use the mechanism in practical applications seems promising, only limited use is predicted as compared to the (room temperature !) Schottky barrier diode or the superconducting devices. The estimated frequency operating range is inferior to the MIM diode.

Acknowledgement.

We are most grateful to Dr. A.G.M. Jansen for helpful discussions.

REFERENCES

1. For recent reviews see: K.M. Evenson and F.R. Petersen in: *Laser Spectroscopy of Atoms and Molecules*, edited by H. Walther (Springer, Berlin, 1976); J.J. Jimenez and F.R. Petersen, Infrared Phys. 17, 541 (1977).
2. For a comparative review see: D.J.E. Knight and P.T. Woods, J. Phys. E: Sci. Instr. 9, 898 (1976).
3. L.M. Matarrese and K.M. Evenson, Appl. Phys. Lett. 17, 8 (1970); B.L. Twu and S.E. Schwarz, Appl. Phys. Lett. 26, 672 (1975).
4. S.M. Faris, T.K. Gustafson and J.C. Wiesner, IEEE J. Quantum Electron. QE-9, 737 (1973).
5. B.L. Twu and S.E. Schwarz, Appl. Phys. Lett. 25, 595 (1974).
6. V. Daneu, D. Sokoloff, A. Sanchez and A. Javan, Appl. Phys. Lett. 15, 398 (1968).
7. A. Sanchez, S.K. Singh and A. Javan, Appl. Phys. Lett. 21, 240 (1972); T.K. Gustafson and T.J. Bridges, Appl. Phys. Lett. 25, 56 (1974).
8. G.M. Elchinger, A Sanchez, C.F. Davies,Jr. and A. Javan, J. Appl. Phys. 47, 591 (1976).
9. H.D. Riccius, Infrared Phys. 17, 245 (1977); H.D. Riccius, Appl. Phys. 24, 215 (1981).
10. N.M. Miskovsky, P.H. Cutler, T.E. Feuchtwang and A.A. Lucas, Int. J. Infrared Millimeter Waves 2, 739 (1981) and references herein.
11. H.-U. Daniel, M. Steiner and H. Walther, Appl. Phys. 25, 7 (1981).
12. D.P. Siu and T.K. Gustafson, Appl. Phys. Lett. 31, 71 (1977); D.P. Siu, R.K. Jain and T.K. Gustafson, Appl. Phys. Lett. 28, 407 (1976).
13. N.M. Miskovsky, P.H. Cutler, T.E. Feuchtwang and A.A. Lucas, Appl. Phys. A 27, 139 (1982).
14. I.K. Yanson, Zh. Eksp. Teor. Fiz. 66, 1035 (1974) /Sov. Phys. JETP 39, 506 (1974)/; A.G.M. Jansen, F.M. Mueller and P. Wyder, Phys. Rev. B 16, 1325 (1977).
15. R.W. van der Heijden, A.G.M. Jansen, J.H.M. Stoelinga, H.M. Swartjes and P. Wyder, Appl. Phys. Lett. 37, 245 (1980).
16. H.J.A. Bluyssen and A.F. van Etteger, unpublished.
17. A. Sanchez, C.F. Davis,Jr., K.C. Liu and A. Javan, J. Appl. Phys. 49, 5270 (1978).
18. C.C. Bradley and G.J. Edwards, IEEE J. Quantum Electron. QE-9, 548 (1973).
19. Y. Yasuoka, T. Sakurada, D.P. Siu and T.K. Gustafson, J. Appl. Phys. 50, 5860 (1979).

20. S. Ramo, J.R. Whinnery and T. van Duzer, *Fields and Waves in Communication Engineering* (Wiley, New York, 1965).

21. H. Jasik, ed. *Antenna Engineering Handbook* (McGraw-Hill, New York, 1961).

22. S. Wang, Appl. Phys. Lett. 28, 303 (1976).

23. H.D. Riccius, Appl. Phys. Lett. 27, 232 (1975).

24. H.M. Swartjes, Doctoral Thesis, Univ. of Nijmegen, 1981.

25. A.R. Kerr, IEEE Trans. Microwave Theory Tech. MTT-23, 781 (1975).

26. D.A. Weitz, W.J. Skocpol and M. Tinkham, J. Appl. Phys. 49, 4873 (1978).

27. H.C. Torrey and C.A. Whitmer, *Crystal Rectifiers* (Massachusetts Institute of Technology Radiation Laboratory Series, Vol. 15, McGraw-Hill, New York, 1948).

28. A.A. Lucas and P.H. Cutler, Solid State Commun. 13, 361 (1973).

29. N.M. Miskovsky, S.J. Shepherd, P.H. Cutler, T.E, Sullivan and A.A. Lucas, Appl. Phys. Lett. 35, 560 (1979).

30. For a review see: A.G.M. Jansen, A.P. van Gelder and P. Wyder, J. Phys. C: Solid State Phys. 13, 6073 (1980); A.G.M. Jansen, Thesis, Univ. of Nijmegen, 1980.

31. Yu. V. Sharvin, Zh. Eksp. Teor. Fiz. 48, 984 (1965) /Sov. Phys. JETP 21, 655 (1965)/

32. G. Wexler, Proc. Phys. Soc. London 89, 927 (1966).

33. I.O. Kulik, A.N. Omel'yanchuk and R.I. Shekhter, Fiz. Nizk. Temp. 3, 1543 (1977) /Sov. J. Low Temp. Phys. 3, 740 (1977)/; A.P. van Gelder, Solid State Commun. 25, 1097 (1978).

34. M. Tinkham, M. Octavio and W.J. Skocpol, J. Appl. Phys. 48, 1311 (1977).

35. C.B. Duke, *Tunneling in Solids* (Academic, New York, 1969), eq. 7.20.

36. A.G.M. Jansen, A.P. van Gelder, P. Wyder and S. Strässler, J. Phys. F: Metal Phys. 11, L15 (1981).

37. E.L. Wolf in: *Inelastic Electron Tunneling Spectroscopy*, edited by T. Wolfram (Springer, Berlin, 1978).

38. A.P. van Gelder, A.G.M. Jansen and P. Wyder, Phys. Rev. B 22, 1515 (1980).

39. J.G. Simmons, J. Appl. Phys. 34, 2581 (1963); S.P. Kwok, G.I. Haddad and G. Lobov, J. Appl. Phys. 42, 554 (1971); S.M. Faris and T.K. Gustafson, Appl. Phys. Lett. 25, 544 (1974).

40. T.K. Gustafson, R.V. Schmidt and J.R. Perucca, Appl. Phys. Lett. 24, 620 (1974).

41. M. Heiblum, S. Wang, J.R. Whinnery and T.K. Gustafson, IEEE J. Quantum Electron. QE-9, 548 (1973).

42. A. Nagasima and T. Tako, Jpn. J. Appl. Phys. 16, 1349 (1977).

43. S.M. Faris, B. Fan and T.K. Gustafson, Appl. Phys. Lett. 27, 629 (1975).

44. B. Fan, S.M. Faris, T.K. Gustafson and T.J. Bridges, Appl. Phys. Lett. 30, 177 (1977).

45. S.I. Green, P.D. Coleman and J.R. Baird, presented at the Symposion on Submillimeter Waves, Polytechnic Institute of New York, Brooklyn, 1970 pp. 369-389.

46. The absolute sign of the polarity should not be compared because it depends on definition. For a (dc) current biased contact, with an I(V) characteristic having a curvature consistent with Fig. 2, the induced voltage is of the same sign as the dc applied voltage. The induced dc current subtracts from the applied current.

47. S.M. Faris and T.K. Gustafson, IEEE J. Quantum Electron QE-8, 558 (1972).

48. A.G.M. Jansen, Doctoral Thesis, Univ. of Nijmegen, 1976; J. Caro, R. Coehoorn and D.G. de Groot, Solid State Commun. 39, 267 (1981).

49. B.I. Verkin, I.K. Yanson, I.O. Kulik, O.I. Shklyarevski, Solid State Commun. 30, 215 (1978).

50. K.C. Liu, C. Davis,Jr. and A. Javan, Phys. Rev. Lett. 43, 785 (1979); see also: T.E. Feuchtwang, P.H. Cutler and N.M. Miskovsky, Phys. Rev. B 23, 3563 (1981).

51. At these high frequencies, in fact a quantum treatment of rectification should be employed: J.R. Tucker and M.F. Millea, Appl. Phys. Lett. 33, 611 (1978). Quantum effects already start to play a role at frequencies of a few THz (see ref. 24).

52. H.R. Fetterman, B.J. Clifton, P.E. Tannenwald and C.D. Parker, Appl. Phys. Lett. 24, 70 (1974); D.T. Hodges and M. McColl, Appl. Phys. Lett. 30, 5 (1977).

53. B.F.J. Zuidberg and A. Dymanus, Appl. Phys. Lett. 29, 643 (1976); D.D. Bicanic, B.F.J. Zuidberg and A. Dymanus, Appl. Phys. Lett. 32, 367 (1978).

54. D. Tsang and S.E. Schwarz, Appl. Phys. Lett. 30, 263 (1977).

55. M. McColl, IEEE Trans. Microwave Theory Tech. MTT-23, 781 (1977); W.M. Kelly and G.T. Wrixon, IEEE Trans. Microwave Theory Tech. MTT-27, 665 (1979).

56. P.L. Richards in: *Semiconductors and Semimetals*, Vol. 12, R.K. Willardson and A.C. Beer, eds., (Academic, New York, 1977) pp. 395-439.

Solid State Communications, Vol. 39, pp. 133–137
Pergamon Press Ltd. 1981. Printed in Great Britain.

0038–1098/81/010133–05$02.00/0

V 1. PHOTON ASSISTED TUNNELING IN SUPERCONDUCTOR–NORMAL METAL POINT CONTACTS AT FAR-INFRARED FREQUENCIES

R. W. van der Heijden, J. H. M. Stoelinga, H. M. Swartjes and P. Wyder

Research Institute for Materials, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands

(Received 3 October 1980, *in revised form* 22 *February* 1981 *by A. R. Miedema)*

The response of normal metal–superconductor point contacts to radiation at frequencies up to 2.5 THz is studied experimentally. The results can be analyzed in terms of the so called Photon Assisted Tunneling (PAT) effect and are in excellent agreement with recent theoretical predictions.

INVESTIGATIONS of the electrical and electrodynamical properties of contacts between two bulk superconductors are very important for the understanding of a variety of superconductive phenomena. Correspondingly, an extensive literature exists on the influence of radiation on superconducting tunnel junctions, point contacts between superconductors, microbridges, and many other types of weak links. However, the normal metal–superconductor (NS) point contact has received relatively little attention, although it is likely to be a very sensitive probe for studying the conversion process of a normal current into a super-current. Only very recently, some progress has been made in order to understand the d.c. $I-V$ characteristics of such diodes in some detail [1–3].

In this paper, we report the first results of an investigation of the high frequency response of an NS point contact. This diode proved to be a sensitive radiation detector whose properties can excellently be described in terms of the well known "Photon Assisted Tunneling" (PAT)-effect, the quantum mechanical analogue of classical rectification. PAT was first observed by Dayem and Martin [4] at lower frequencies and has been explained subsequently by Tien and Gordon [5]. The effect has since been confirmed in a number of experiments in superconductor–insulator–superconductor tunnel junctions [6]. PAT has also been observed in superconducting thin film weak links [7] and superconducting point contacts [8]. Point contacts are particularly well suited for analysing the effect as their small contact area favours a good comparison with theory [6] and minimizes shunt capacitance, allowing the use of much higher frequencies. Note that in the former devices, the PAT-effect arises because of the nonlinearities in the $I-V$ characteristic caused by the quasiparticle current onset at the voltage corresponding to the energy gap of the superconductor. Recently, the theory of the PAT effect has been generalized and extended [9] in view of the renewed interest in using superconducting tunnel junctions as microwave receivers and mixers [10]. It was shown that the practical merits of superconductors are enhanced by the specific features introduced by PAT [11].

Point contacts between superconductors are either of the Josephson type, if the coupling is strong, or of the quasi particle tunneling type, if the coupling is weak. In the first case, the high frequency response is dominated by the Josephson effect, even for the region of high frequency and high voltage bias and for contacts showing strong gap-related nonlinearities [12]. In the latter case, they show the PAT-effect related to the nonlinearity in the $I-V$ characteristic due to the energy gap of the superconductor. For contacts between a normal metal and a superconductor, as investigated in the present study, the Josephson effect can be ruled out [13]. Furthermore, for the moderately low contact resistances used ($\lesssim 100\,\Omega$), no insulating oxide layer exists at the interface and therefore quasi-particle tunneling is not very important in these contacts. The nonlinearity of the $I-V$ characteristic is then due to the transformation of a normal current into a supercurrent and is so large, that it is to be expected that the influence of the radiation at the frequencies used can be described in a phenomenological way by the PAT effect.

For the present experiments, contacts were made at liquid helium temperatures by means of a differential screw mechanism between a sharpened 50 μm diameter Cu wire and a superconducting post consisting of either Nb or Sn. The junction was mounted at the end of a copper cone terminating a stainless steel lightpipe with a diameter of 12 mm, with the copper wire of the point contact perpendicular to the axis of the cone. Current–voltage characteristics $I(V)$ as well as the first derivatives dV/dI vs voltage V were recorded, using standard current modulation techniques. High frequency effects were studied by illuminating the point contact with radiation at three typical far-infrared frequencies, generated by an optically pumped far-infrared laser: 246 GHz

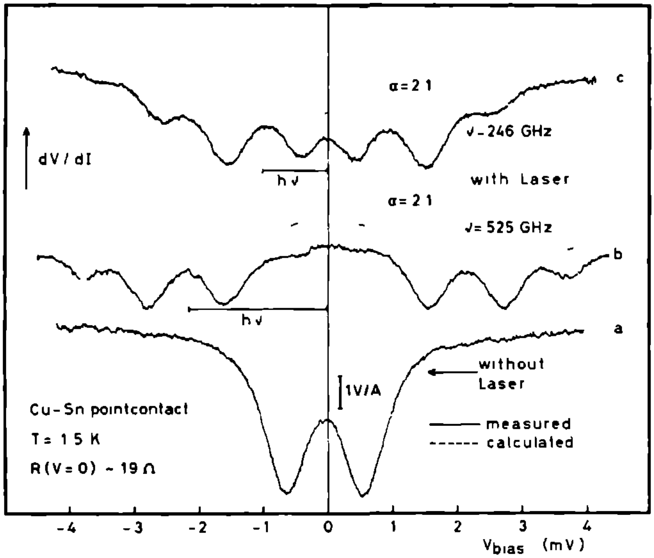


Fig. 1. Resistance vs voltage for a Sn–Cu point contact at $T = 1.5$ K. $R(V=0) \simeq 19\,\Omega$. (a) without radiation. (b, c) drawn lines: same as a) but now irradiated with frequencies 525 and 246 GHz respectively; dashed lines: curves calculated from a) according to PAT model. For both curves a value of $\alpha = 2.1$ was used. Horizontal bars indicate the relevant photon energy.

($\lambda = 1.22$ mm), 525 GHz ($\lambda = 571\,\mu$m) and 2.52 THz ($\lambda = 119\,\mu$m) [14].

A typical differential resistance dV/dI vs. voltage characteristic without the far infrared radiation is shown in Fig. 1(a). The overall shape is in good agreement with the results found by Gubankov *et al.* [1]. After an initial decrease of the resistance as a function of the bias voltage, the resistance starts to increase at a voltage corresponding to the energy gap $\Delta(\Delta \sim 0.6$ mV for Sn), for voltages well in excess of Δ, the resistance is approximately constant. It is typical for the SN contact that the value of the resistance at $eV \gg \Delta$ is higher than for $eV \lesssim \Delta$, as opposed to the situation for a simple tunneling like behaviour.

Theoretically, the d.c. behaviour of the SN point contact was studied recently by Artemenko *et al.* [2] and by Zaitsev [3]. Qualitatively, the I–V characteristic can be understood as resulting from two different effects associated with the current transport through a normal metal–superconductor interface, namely the conversion of a "normal" current into a "super current" of pairs and a current due to quasi particles. The conversion mechanism is analogous to the Andreev scattering process. Artemenko *et al.* [2] and Zaitsev [3] have found that for small voltages ($eV \lesssim \Delta$) the conversion current contribution increases linearly with V and then becomes independent of voltage for voltages well in excess of the gap voltage as then the maximum number of electrons is converted into pairs. The additional current is carried by single electrons (or quasi particles). This saturating SN conversion current leads to the typical "excess current" usually observed in SN point contacts, i.e. the high voltage portion of the I–V characteristic does not simply extrapolate back linearly to the origin. The conversion current leads to a finite (relatively low) resistance at arbitrarily low voltages, whereas the saturation of the SN conversion current reveals itself in a resistance rise in the R–V characteristic at higher voltages ($eV > \Delta$) [Fig. 1(a)]. Parallel to the conversion current, there is also a current contribution from electrons being transferred between single electron and quasi particle states in the normal metal and superconductor respectively, in a way analogous to tunnel junctions (superconductor–insulator–superconductor, or superconductor–insulator–normal metal configuration). The initial decrease of the resistance up to $eV = \Delta$ is a manifestation of this single particle current as single particle transfer is less effective at $eV \lesssim \Delta$. This mechanism is responsible for the structure often seen at $eV = 2\Delta$ in the I–V characteristics of superconducting point contacts [8, 15] and in thin film weak links [7]. It should be noted that the relative maximum around zero voltage was not present in all contacts investigated; sometimes only a broad resistance minimum was

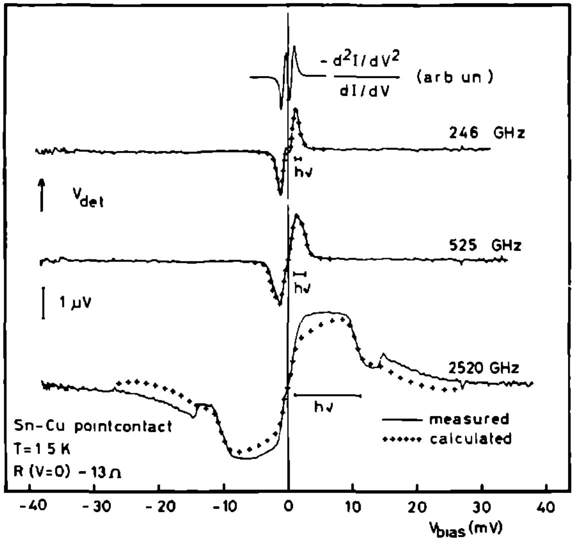


Fig 2 Upper curve $-(d^2I/dV^2)/(dI/dV)$ vs voltage V, calculated from measured dV/dI vs V curve (not shown, but very similar to Fig 1(a)) for Sn–Cu point contact at $T = 1\ 5$ K $R(V = 0) \simeq 13\,\Omega$ Lower three curves drawn lines, phase sensitively detected voltage across the same junction when irradiated with chopped laser beam (frequencies 246, 525 and 2520 GHz respectively) using a chopping frequency of 560 Hz, as a function of bias voltage The laser power was kept sufficiently low in order to keep the shape of the curves independent of power Crosses calculated responsivities from dV/dI vs V curve according to the PAT model in the small signal limit, the amplitude is scaled such as to give good comparison Horizontal bars indicate photon energy in units of electron volts

observed This can be understood on the basis of the analysis of Zaitsev [3] who points out that the relative maximum should only be present in the Knudsen limit, i e if $l < a$, where l is the mean free path of the electrons and a the diameter of the contact

Usually, we also observed an apparent quenching of superconductivity in the contact region, which manifested itself either by a small and slightly hysteretic discontinuous voltage step in the $I-V$ characteristic or as a rather narrow smooth resistance peak This effect has been observed in such contacts more often [16] and can be explained by the combined effects of critical current density and Joule heating in the superconducting electrode [16] However, for the present analysis we used only $I-V$ characteristics where such quenching effects occurred well outside the voltage range of interest

Figures 1(b) and 1(c) show the dV/dI vs V characteristic of the NS contact when irradiated with 525 GHz ($\hbar\omega = 2\ 17$ meV) and 246 GHz ($\hbar\omega = 1\ 02$ meV) radiation respectively In a qualitative way, it looks from Fig 1(b) as if the original dV/dI characteristic without the radiation is shifted on the voltage scale by an amount of $\hbar\omega/e$ with respect to $V = 0$ This has been substantiated by the theoretical analysis of Artemenko *et al* [2] and Zaitsev [3] who showed that the effects of microwave irradiation on the d c $I-V$ characteristics of the SN point contact can be described in a way analogous to the PAT formalism the $I-V$ characteristic under the influence of radiation can be expressed simply in terms of the $I-V$ characteristic without radiation by [5, 9]

$$I^{rad}_{dc}(V) = \sum_{n=-\infty}^{\infty} J_n^2(\alpha) I_{dc}\left[V + n(\hbar\omega/e)\right] \quad (1)$$

Here, $I_{dc}(V)$ is the d c current as a function of applied d c bias voltage V without radiation, and J_n represents the ordinary Bessel function of order n In the derivation of equation (1) it is assumed that the voltage across the contact is modulated as $V_1 \cos \omega t$, and $\alpha = (eV_1/\hbar\omega)$, where ω is the angular frequency of the radiation For better data handling the measured dV/dI vs V curve (curve 1^a) was digitized and stored in a computer Then dV/dI with radiation applied was calculated, starting from the measured dV/dI without radiation and using equation (1) The curves calculated in this way are shown as dashed lines in Fig 1(b) and 1(c) The value of the parameter α, a measure of the power actually coupled into the junction, was adjusted to the experimental curves, but no attempt was made for a detailed fit As can be seen from Fig 1, a very satisfactory agreement can be obtained between the measured and calculated characteristics of the diode under irradiation

Figure 2 shows the video responsivity for three different frequencies of a very similar contact as used for Fig. 1. In these experiments, the laser is chopped at 560 Hz and the voltage across the junction is detected phase sensitively as a function of the bias voltage. Also in this mode of operation, the quantum mechanical detection mechanism is evident most notably at the highest frequency the width of the sensitive region scales with the photon energy $h\nu$. The curves in Fig. 2 were obtained at strongly reduced power levels, such that their shape was almost independent of the power which allows comparison with theory in the small signal limit. Again, for comparison, the responsivity is calculated, starting from the measured dV/dI curve without radiation [9]; the calculated values are given as crosses in Fig. 2. The characteristic features of the experimental findings are reproduced very well, although significant differences do exist at 2 5 THz Note that a more quantitative comparison should also take into account the impedance mismatch variation as a function of bias voltage, which is neglected here. In addition, a rather fundamental limitation of the procedure sketched above is the fact that in the calculations a high frequency voltage bias across the contact was assumed, obviously, for the rather low Ohmic junctions used in the present study, a current bias would be more appropriate. It is of some interest to compare the video responsivity with classical rectification For that purpose, Fig. 2 shows also the simple $(d^2I/dV^2)/(dI/dV)$ signal without radiation, which, at low frequencies, is proportional to the classical responsivity based on rectification As can be seen, the quantum responsivity curve at our lowest frequencies is not too much different from the frequency independent classical result. Based on very general arguments [9, 17], quantum behaviour of a nonlinear element can be expected at frequencies ω whenever the resistance of the element varies rapidly on a voltage scale given by $\hbar\omega/e$. This condition is clearly fulfilled in our case (Figs. 1 and 2). From a slightly different point of view, nonclassical behaviour should occur if the classical value $\frac{1}{2}(d^2I/dV^2)/(dI/dV)$ of the current responsivity (in our experiments typically $\sim 200\,V^{-1}$) becomes of the same order of magnitude as the quantum limited value of one electron per absorbed photon at the frequencies used ($e/\hbar\omega \simeq 200\,V^{-1}$ at 1200 GHz) Note also that our highest frequency corresponds to ten times the energy gap of the superconducting electrode, and the value $\hbar\omega/e$ exceeds the whole region of the nonlinearity at both sides of zero bias by almost an order of magnitude Therefore, the rather unusual and extreme circumstances of the present experiments provide a rather crucial test of the general theory [9].

It should be emphasized that in order to get good agreement between the calculated curves and the experimental ones, the full nonlinearity of the I V characteristic has to be used, including the part attributed to SN conversion. This has very explicitly been seen on some of our contacts where the gap-related structure around $V = \pm\Delta$ was completely absent and the resistance showed only a broad minimum around $V = 0$. In the general theory of photon assisted tunneling [9], it is explicitly assumed that single electrons are transferred between single electron eigenstates in both electrodes. This assumption is obviously valid in the case of the non-linearity caused by quasi-particle current onset at $eV = \Delta$, leading to PAT not only in real tunnel junctions, but also in the case of, e.g. thin film weak links [7, 18]. However, this assumption is less obvious for the process of converting a normal-current into a supercurrent, but our results confirm the theoretical predictions by Artemenko *et al* [2] and by Zaitsev [3].

In conclusion, we have observed the PAT-effect in an SN point contact at the highest frequency up to now, in agreement with general theories on PAT [5, 9] as well as with more specific theories on superconductivity [2, 3]. These results can be taken as an experimental support for the proposed model [2, 3] of current transport in an SN point contact The structure of the $I-V$ characteristic, at least at not too high voltages, is due to really microscopic processes and is not related to macroscopic effects destroying superconductivity, such as heating or critical current density

Acknowledgements – We are most grateful to Dr A.G.M. Jansen for his helpful contribution to this work Part of this work has been supported by the "Stichting voor Fundamental Onderzoek der Materie" (FOM) with financial support of the "Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek" (ZWO).

REFERENCES

1. V.N. Gubankov & N M. Margolin, *Pis'ma Zh. Eksp. Teor. Fiz.* 29, 733 (1979) [*JETP Lett.* 29, 673 (1979)].
2. S.N. Artemenko, A.F. Volkov & A.V. Zaitsev, *Solid State Commun.* 30, 771 (1979)
3. A V. Zaitsev, *Zh. Eksp. Teor Fiz* 78, 221 (1980) [*Sov Phys. JETP* 51, 111 (1980)].
4. A.H. Dayem & R.J. Martin, *Phys Rev Lett.* 8, 246 (1962).
5. P.K Tien & J P Gordon, *Phys Rev* 129, 647 (1963)
6. See, e.g C.A Hamilton & S. Shapiro, *Phys Rev* B2, 4494 (1970)
7. B Kofoed & K. Saermark, *Phys Rev. Lett.* 31, 1124 (1973).
8. H Thome, Y Couder & A. Libchaber, *J. Appl. Phys* 49, 1200 (1978)
9. J R Tucker & M F Millea, *Appl Phys Lett* 33, 611 (1978), J.R. Tucker, *IEEE J Quantum Electron* QE-15, 1234 (1979).

10 P L Richards, T M Shen, R E Harris & F L Lloyd, *Appl Phys Lett* **34**, 345 (1979), G J Dolan, T G Phillips & D P Woody, *Appl Phys Lett* **34**, 347 (1979), S Rudner & T Claeson, Appl Phys Lett **34**, 711 (1979)

11 J R Tucker, *Appl Phys Lett* **36**, 477 (1980), P L Richards, T M Shen, R E Harris & F L Lloyd, *Appl Phys Lett* **36**, 480 (1980), T M Shen, P L Richards, R E Harris & F L Lloyd, *Appl Phys Lett* **36**, 777 (1980)

12 D A Weitz, W J Skocpol & M Tinkham, *Phys Rev Lett* **40**, 253 (1978), D A Weitz, W J Skocpol & M Tinkham, *Phys Rev* **B18**, 3282 (1978)

13 The possible existence of phase slip centers [W J Skocpol, M R Beasley & M Tinkham, *J Low Temp Phys* **16**, 145 (1974)] the immediate contact region cannot be ruled out for the NS point contacts, but no direct experimental evidence could be found

14 J J Gallagher, M D Blue, B Bean & S Perkowitz, *Infrared Phys* **17**, 43 (1977)

15 D A Weitz, W J Skocpol & M Tinkham, *J Appl Phys* **49**, 4873 (1978)

16 O Iwanyshyn & H J T Smith, *Phys Rev* **B6**, 120 (1972)

17 E D Dahlberg, R L Orbach & I Schuller, *J Low Temp Phys* **36**, 367 (1979)

18 In agreement with the assumption in the theoretical models [2, 3], we believe to have real metallic contacts in our experiment This is supported by the fact that stable contacts are usually obtained by a burn in process, and the same assumption is found to describe best similar point contacts between either superconductors [15] or normal metals [A G M Jansen, A P van Gelder & P Wyder, *J Phys C Solid State Phys* **13**, 6073 (1980)]

V 2. HIGH-FREQUENCY PROPERTIES AND CURRENT-VOLTAGE CHARACTERISTICS OF NORMAL METAL-SUPERCONDUCTOR AND NORMAL METAL-NORMAL METAL POINT CONTACTS.

ABSTRACT

Current-voltage characteristics of point contacts between either normal metals (NcN) or between a normal metal and a superconductor (NcS) are strongly nonlinear. The dc-behaviour of an NcN contact is very well understood, while recently there has been great progress in the theoretical understanding of the NcS contact. In this paper, I(V) characteristics of NcS contacts have been investigated under conditions where they display the typical behaviour of an NcN contact when the superconductor is in the normal state.

For the NcS contacts, both the direct detection as well as the I(V) characteristic under application of radiation are clearly dominated by quantummechanical effects (the so called Photon Assisted Tunneling (PAT) effect), because the I(V) characteristic is nonlinear on a voltage scale much smaller than $\hbar\omega/e$. ($\hbar\omega/e$ is the far-infrared photon energy in units of voltage.)

For the NcN contacts, the nonlinearity spans a voltage range at least an order of magnitude larger (determined here by the Debye energy instead of the superconducting energy gap with NcS contacts). However, also in this case, at the highest far-infrared frequencies, the experimental results strongly indicate that the PAT-effect is present here as well.

2. 1. Introduction.

The method of forming very small electrical contacts between two bulk metals by use of the point contact technique at low temperatures is very well established, not only as a tool for investigating a variety of fundamental physical phenomena, but also because it offers a potential interesting device for practical applications, most notably in high-frequency operations. A main virtue of the point contact is the ability to have a perfect or nearly perfect metallic contact between the electrodes, so that electronic conduction in both electrodes is strongly coupled (by a "constriction"), as opposed to tunneling devices where both electrodes are largely decoupled. The oldest, most widely known and intensively investigated point contact is that with two superconducting electrodes (ScS). The Josephson effect is the most prominent feature of the ScS contact and widely exploited and appreciated in high-frequency operation[1-3]. Also the (constriction type) point contact between normal metals (NcN) is by now well established as a tool to investigate some fundamental properties of metals[4]. Recently, an attempt has been made to correlate some specific features of the ScS contact, in addition to the Josephson effect, with typical characteristics of the NcN contact[5]. While the NcN contact appears to be well understood theoretically[4], this was until recently not the case with the ScS contact. Recent theoretical work[6-8] not only clarified several previously ambiguously understood properties of the ScS contact, but also revealed a connection between the ScS contact and the normal metal-superconductor point contact (NcS). Experimentally, a renewed interest for the NcS systems also exists[9-11].

High-frequency investigations were reported for the NcN contact in the regime of classical rectification[12,13] and for arbitrary frequencies for the NcS contact[14]. The properties probed in the NcN-contact experiments (Eliashberg function)[4], even for typically nonsuperconducting metals such as the noble metals (e.g. Cu), are strongly connected with the theories and properties of superconductivity[4]. Investigations of ScS and NcN contacts are useful to study superconductivity or normal metallic behaviour. It is of interest to investigate whether mixed experiments of this kind may contribute towards a better and more unified understanding of metals and their superconducting state. The parallel studies of ScS and NcN contacts reported in ref. 5 in principle do serve this purpose, while the NcS system is relevant in this respect as well.

In this work, the effects of high-frequency irradiation of ScN contacts[14] will be reported and discussed in more detail. Because the high-frequency effects are intimately connected with the shape and origin of the current-voltage characteristics (I(V)), these will be discussed to some extent separately and compared with the theory, as well as with other published data, to characterize the contacts. Some high-frequency data will be given for NcN contacts also, at higher frequencies than before[12,13]. The experimental data strongly suggest that the so called "Photon Assisted Tunneling" model (PAT), which applies to the ScN contacts[14], holds for the NcN contact as well. This theory was originally derived for superconducting tunnel junctions[15], but applies, in fact, to every single electron tunneling mechanism[16]. It was explicitly derived for the ScN contact also[6,7].

The experimental set up used for the present experiments is the same as that used previously[12-14]. In short, first and second derivatives of the I(V) characteristics are recorded, using standard current modulation techniques. Radiation is obtained from an optically pumped far-infrared laser. The contact is mounted at the end of a lightpipe which is illuminated by the laser. Contacts are made and adjusted at liquid helium temperatures. In all cases reported here, the point was made from 50 μm diameter Cu wire, etched to a sharp point. The post consisted either of the superconductors Nb or Sn in the ScN case or of Cu in the NcN case . Apart from etching, the contacting surface was not especially treated.

2. 2. Current-voltage characteristics.

Figure 1 shows two representative results of the differential resistance R_d ($\equiv dV/dI$) versus applied bias voltage curves for an NcN and ScN contact respectively. As can be seen, the overall shape of both curves is very similar, but the physical origin is entirely different in the two cases. The result for the NcN case can be explained by assuming that the contact is in the clean limit, i.e. $\ell \gg a$ where ℓ is the electron mean free path and the electrodes are assumed to be in metallic contact (that is no interface layer) over an infinitely thin orifice of radius a[4]. Under this condition, electrons are ballistically injected from one electrode into the other, with an increase in energy of eV (e electron charge, V voltage across contact). For ballistic transport the current is only proportional to the area of the contact, the electron density and the Fermi velocity, but does not depend on the nature of

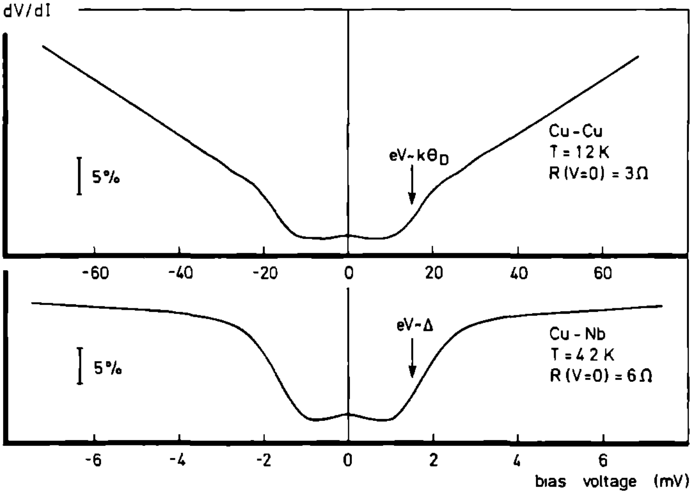


Fig. 1. Typical resistance vs voltage characteristics for point contacts between normal metals (upper trace) and between a normal metal and a superconductor (lower trace). The point of maximum inflection is either dictated by the Debye energy ($k\Theta_D$) or the superconducting energy gap (half gap = Δ).

the material, notably scattering. Neglecting geometrical factors:

$$I = e \cdot \Delta n \cdot v_F a^2 = e(\frac{dn}{dE}/_{E_F} \cdot eV) \cdot v_F a^2$$

so that the resistance becomes[4]:

$$R_{Sh} \sim \frac{1}{(\frac{dn}{dE}/E_F)e^2 v_F a^2} \sim \frac{\rho \ell}{a^2} \qquad (1)$$

which is the so called Sharvin resistance. Here Δn is the electron density passing the hole, v_F the Fermi velocity, $\frac{dn}{dE}/E_F$ the electron density of states at the Fermi level and V the voltage across the contact, ρ the resistivity of the metal, $\rho\ell = m^* v_F/ne^2$ with m^* electron effective mass and n total electron density. After injection, scattering in the bulk material contributes to the total resistance by an amount given by the spreading resistance R_s:

$$R_s = \frac{\rho}{2a} \qquad (2)$$

which is proportional to the bulk resistivity ($R_s << R_{Sh}$ for $\ell >> a$). When the electrons are accelerated to the energy of the zone boundary accoustic phonons, the scattering rate is increased (increase in R_s) and this leads to the increase in resistance at $eV \sim k\Theta_D$ where $k\Theta_D$ is the Debye energy[4] (∿ 17 meV for Cu) (see Fig. 1). The (nearly linear) increase of $R_d(V)$ for $eV > k\Theta_D$ results from the so called "background"[4] and is attributed to a nonequilibrium phonon distribution near the orifice[4]. It is not essential for the present work.

From the theories[6-8], it follows that the I(V) for a ScN contact can be understood in terms of Andreev scattering processes at the normal metal-superconductor interface. An electron starting from N recombines with another electron at the interface to form a Cooper pair in S, leaving an empty electron state below the Fermi level in N (hole). Thus at the interface electrons are (antispecularly) reflected as holes, as long as $eV < \Delta$ where Δ is half the superconducting energy gap. For $eV > \Delta$ the differential resistance is the same as if both electrodes were normal and given by the resistance for ballistic transport through a hole, R_{Sh} (neglecting the spreading resistance R_s). For $eV < \Delta$ this resistance is theoretically reduced by a factor of 2 [17], because each electron has a double contribution to the current (it flows as a Cooper pair in S and has reflected a hole in opposite direction in N). Thus $R_d(V=0)$ for an ScN contact is just the average of $R_d(V=0)$ for ScS and $R_d(V=0)$ for NcN. Although these qualitative arguments essentially describe the process, the experimental results are somewhat more complicated and will be discussed in more detail. (Note e.g. that the resistance change at $eV \sim \Delta$ is much less than a factor 2.)

Clearly, the similarity between the two curves in Fig. 1 is completely accidental and uncorrelated. An important difference is the voltage range where the resistance variations occur . This range is set by $k\Theta_D$ for NcN ($\sim$ 17 meV for Cu) and by Δ for ScN (1.5 meV for Nb and 0.6 meV for Sn) and thus differs by more than an order of magnitude. Note also the principal difference in the physical processes. For NcN, at $eV \gtrsim k\Theta_D$ the resistance rises above R_{Sh} because of the increased scattering of normal electrons which, in fact, scatter back through the hole[4]. For ScN at $eV \lesssim \Delta$, the resistance decreases below R_{Sh}, because in addition to each transmitted electron, a hole is reflected. The detailed comparison of the two curves as in Fig. 1, however, will turn out to be instructive in connection with the high-frequency behaviour in both contacts.

It is important to stress that the $R_d(V)$ curve for an ScN contact essentially differs from that of a tunnel junction between a normal metal and a superconductor (SIN). In the latter case, current flows only (at T=0) when the quasiparticles are elevated above the energy gap Δ. Thus for an SIN contact R_d is very large for $eV < \Delta$ (∞ at T=0 and V=0, ideally) and reaches a finite, normal state value for $eV > \Delta$. The I(V) curve for an ScN contact as in Fig. 1 is, for $eV > \Delta$ (and $kT \ll \Delta$), usually written as:

$$I(V) = I_{exc} + \frac{V}{R_{Sh}} \tag{3}$$

where I_{exc} is the so called excess current and expresses the fact that the I(V) characteristic at high voltage ($eV > \Delta$), although linear, does not extrapolate back through the origin. The excess current may be viewed as the extra current which flows as a result of Andreev reflection. As compared to the NcN case, transmitted electrons with energy ε in the range $0 < \varepsilon < \Delta$ produce an Andreev reflected hole in N (or Cooper pair in S), so that roughly $I_{exc} = \frac{\Delta/e}{R_{Sh}}$, independent of voltage. The theory[8] gives:

$$I_{exc} = \frac{4}{3}\,\frac{\Delta}{eR_{Sh}} \quad (eV \gg \Delta,\ kT \ll \Delta) \ . \tag{4}$$

As an interesting peculiarity, it may be noted that even for the NcN case, the I(V) curve in a certain voltage range for $eV > k\Theta_D$ can be parametrized with an expression analogous to eq. (3). Formally, it might thus be stated that all (pure) constriction type contacts, at low temperature, whether superconducting or not, are characterized by a phenomenon of "excess current" in some voltage range. For the NcN contact however, I_{exc} has no simple physical meaning. Moreover, when compared to the Ohmic curve $I = V/R_{Sh}$ as a reference, the NcN contact is more naturally understood in terms of a (voltage dependent) deficient current for $eV > k\Theta_D$, rather than an excess current at low voltage.

The experimental curves however deviate from the simple situation sketched qualitatively above, while the theory also is much more detailed. For the NcN contacts, the main observations are well reproduced by the theory[4]; the data do vary somewhat from contact to contact as far as the magnitude of the background (rise of $R_d(V)$ with V for $eV > k\Theta_D$) is concerned[4] and in the occurrence of zero bias anomalies[4]. The latter are usually much smaller than resistance changes at $eV \gtrsim k\Theta_D$. The scatter in the shapes of I(V) characteristics for ScN contacts observed for different junctions, appears to be much larger. Analyzing the $R_d(V)$ curves obtained on a great number of different contacts, it appeared useful to classify them into 4 categories, for each of which a representative example is given in Fig. 2. The most general and frequently observed curve has the W shape as in curve I. In addition to the W-shape there is usually also a sharp resistance peak at voltages outside the W-range (curve II). This peak varies strongly from contact to contact in width, magnitude as well as position. In a minority of contacts only a broad minimum around V=0 was observed (curve III). The peaks of curve II are, in fact, rather general and occur also for types I and III, but usually in a much higher voltage range. In this respect, type II is not an essentially different category, but differs from I or III only in that the resistance peaks occur at a very low voltage.

The inset in Fig. 2 shows an I(V) curve of a contact with $R_d(V)$ curve of type IV, but taken on a different junction. A glance at Fig. 2 might suggest that type IV simply is a combination of type III, with the peaks of type II close to V=0. This view is supported by the observation that a relative resistance maximum at V=0, reminiscent to the relative maximum at V=0 in type I or II, sometimes also is present in type IV (see e.g. Fig. 6). It will be shown that types I, II and III can be explained within the framework of the microscopic theories of refs. 6-8, but despite its similarity,

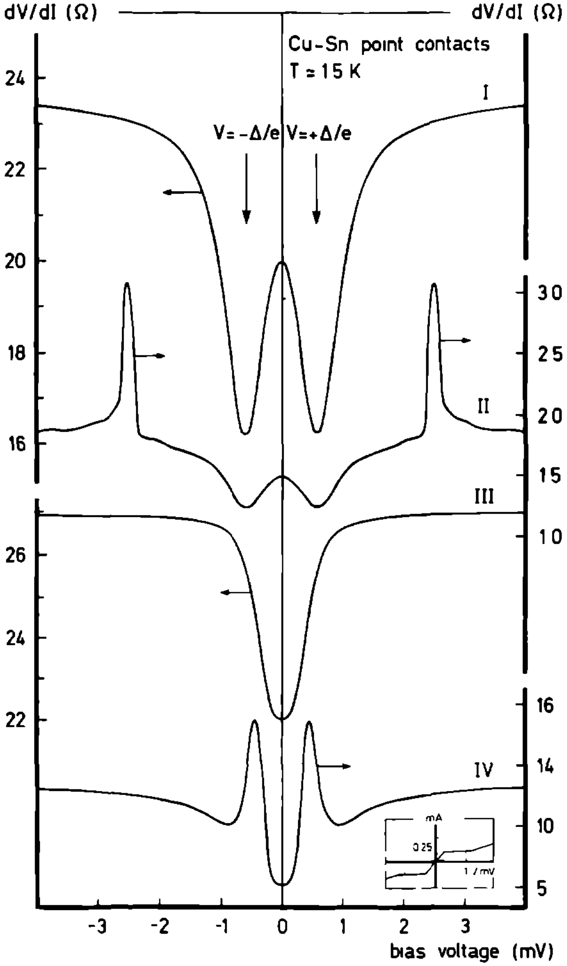


Fig. 2. Types of I(V) characteristics observed for superconductor-normal metal point contacts. Type I is the most commonly observed and is the most general one as expected from the theory. Type II is characterized by a sharp resistance peak corresponding to quenching of the superconductor. Type III is a special form of type I where the relative maximum near zero voltage is missing. Type IV differs essentially from types I-III The inset shows the shape of the current-voltage characteristic yielding a type IV dV/dI vs V curve (but taken from a different contact as for curve IV).

type IV is believed to be essentially different from I, II and III.

2. 2 i. Nonlinearities due to Andreev scattering.

In the theoretical analysis of the ScN point contact problem, two different approaches exist. A Green function formalism is used by Artemenko et al.[6] and Zaitsev[7]. The coherence length $\xi(0)$ is always assumed large compared to the constriction dimensions (length and diameter). I(V) characteristics were then calculated in the dirty[6] ($a >> \ell$) and in the pure[7] ($a << \ell$) limit. In the pure limit, I(V) characteristics similar to those labeled type III are obtained, as discussed qualitatively above.

Apart from the typical shape of the I(V) characteristic, another typical feature of such contacts is the temperature dependence at zero bias. For $\ell >> a$, $r \equiv R_d(V=0)$ is a monotonous function of temperature, starting from $r = R/2$ at $T=0$ to $r=R$ at $T=T_c$, where R is the normal state resistance (or the resistance at high bias voltage). The result for the dirty limit[6] is that $R_d(V=0)$ equals R, then decreases to a minimum at $V \sim \pm \Delta/e$, after which $R_d(V=0)$ rises again to a value R at $V \gtrsim \Delta/e$ This behaviour is qualitatively displayed by the contacts of type I. Experimentally, however, $R_d(V=0)$ is generally smaller than the predicted value R. The physical origin for the relative resistance maximum at $V=0$ is a process analogous to quasiparticle tunneling; single quasiparticle tunneling is not very efficient at $V < \Delta/e$, so thatthe resistance decreases with voltage. Because the tunneling resistance decreases with temperature, as opposed to the Andreev scattering contact resistance, the resistance $R_d(V=0)$ shows a nonmonotonous temperature dependence. Being equal to R at $T=0$ and $T \geqslant T_c$, it shows a minimum at $T \sim 0.8\ \Delta/k$[6].

A different approach to the ScN contact problem was adopted by Klapwijk et al.[8]. Current transport through the contact was handled in terms of the Bogoliubov equations, describing the excitations of a BCS superconductor. Ballistic transport through the contact was assumed throughout, i.e. $\ell >> a$. Using these concepts, it was shown[8] that the qualitative effects can be physically understood using a simple semiconductor representation of the superconductor, thereby contributing to a greater physical insight. In addition to the assumption $\ell >> a$, however, an extra parameter was introduced describing some details of the contact. A δ-function potential barrier was assumed at the interface and a contact is characterized by a dimensionless parameter $Z \equiv H/(2E_F/k_F)$ describing the strength of the barrier. H is the magnitude of the δ-function potential ($V(x) = H\delta(x)$), E_F Fermi energy and k_F Fermi momentum. Effects due to a localized increased impurity content at the contacting surfaces, possible deformation of the metals near the contact, some material asymmetry, e.g. differences in Fermi momenta, up to a complete insulating oxide layer may all be taken into account by the barrier strength. Z=0 corresponds to the ideal situation of a pure metallic constriction, whereas Z >> 1 corresponds to a tunnel junction. Using this single parameter, the complete range of junctions, from pure tunneling (with $R_d(V=0) = \infty$ at T=0) to pure metallic could be continuously described. For Z=0 the results of Zaitsev[7] were retrieved (after correcting the excess current term with a factor 2[17]), while also the results of Artemenko et al.[6] may be reproduced by chosing a suitable value of Z[8].

Published experimental results have shown that the theories of refs. 6 and 7 are not completely adequate to explain the observed behaviour. Very low-Ohmic contacts such as those used by Dorozhkin[10] ($0.1 < R < 1$ Ω) with the estimated values for the contact diameter a ∿ 4000 Å and ℓ ∿ 120 Å, clearly in the dirty limit, do qualitatively show the pure limit behaviour; the I(V) curve is of type III and $R_d(V=0)$ varies monotonously with temperature. Still lower Ohmic contacts, R < 0.1 Ω[18], also have a monotonous $R_d(V=0, T)$ behaviour, although it is not clear whether the other conditions of the theory were fulfilled for these contacts. On the other hand, very high Ohmic contacts (R ≳ 100 Ω) with typical estimates a ∿ 40 Å[9], where the conditions for the pure limit might be expected to be fulfilled, have a strong tunneling character, expected from the theory for the dirty case. $R_d(V=0)$ is roughly equal to R or even larger[9], while $R_d(V=0, T)$ shows a minimum near $T \sim 0.8 \frac{\Delta}{k}$.

The contacts used for the present work lie in a resistance range intermediate between those of ref. 9 and those of ref. 10 ($5 \lesssim R \lesssim 100$ Ω). The behaviour of the contact resistance as a function of V also is intermediate between the behaviour found in ref. 9 and ref. 10. The strength of the relative maximum in $R_d(V)$ at V=0 varies from contact to contact, but $R_d(V=0)$ is always considerably smaller than R, or even completely absent. Curves I and III in Fig. 2 can be taken as representing the extreme behaviours of $R_d(V)$ for the contacts used. Investigations of R_d at V=0 as a function of temperature were undertaken at a few contacts only (see Fig. 3 for an example). Typically, $R_d(V=0, T)$ is almost, but not completely, monotonous and the behaviour shown in the inset in Fig. 3 is intermediate between the corresponding curves in ref. 9 and 10. There is no obvious correlation between the shape of the $R_d(V)$ curves and the resistance of the contact, in the resistance range considered.

For the present ScN junctions, it was verified for several contacts that they invariably showed the usual nonlinearities at $eV \sim k\Theta_D$[4], when the superconductor was in the normal state. This was easily verified for Sn as the superconductor, but is somewhat more complicated for Nb because there was no provision for accurate temperature control above T_c of Nb (~ 9 K), while superconductivity could not be completely suppressed in moderate magnetic fields (< 3 T). The presence of the normal state nonlinearities strongly supports the assumption of ballistic transport ($\ell >> a$) through a clean orifice[4] and cannot be reconciled with the prerequisites of the theory of Artemenko et al.[6]. It was shown by Klapwijk et al.[8] that the shape of the I(V) characteristic, notably the presence and magnitude of the relative maximum $R_d(V=0)$, strongly depends on the strength of the barrier, Z, especially for small barrier strength. From the experimental observation that $R_d(V=0)$ generally is much less than the normal state resistance (or $R_d(V >> \Delta)$), it can be inferred from the model calculations at T=0 of ref. 8, that the parameter Z takes on values $Z \lesssim< 0.5$. Z should in fact be close to 0 for contacts of type III. These values of Z correspond to transmission probabilities for normal electrons $D = 1/(1+Z^2) \gtrsim> 0.8$. The presence of a barrier at the interface with such high transmission probabilities hardly affects the nonlinearities of an NcN contact: the barrier only reduces the strength of the nonlinearity $(1/R)(dR/dV)$ by a factor $D^{\frac{1}{2}}$ ($\gtrsim> 0.9$) as compared to a barrierless case[4]. These high transmission probabilities correspond to only slight deviations from the ideal pure metallic constriction. These small deviations however are random and may vary from contact to contact, thus explaining the rather arbitrary variations of the $R_d(V)$ characteristics observed, especially as far as the structure near V=0 is concerned.

The fact that type III contacts (i.e. without observable maximum in $R_d(V)$) are obtained only in a minority of cases, suggests that highly ideal contacts are somewhat difficult to obtain. A detailed comparison of experimental data with the theory of ref. 8 was recently made[11] for point contacts between Nb and Cu. The experimental results could be excellently described by the theory[8], if Z was adjusted, but the lowest Z-value observed was ~ 0.3. It was suggested[11] that the lower bound on Z resulted from the different values of the Fermi velocities for the metals Nb and Cu.

The consistency of the results for the NcN and ScN contacts within the framework of the theories of refs. 4 and 8 for both diodes, as well as the limited applicability of the theories of refs. 6 and 7 shows that the theo-

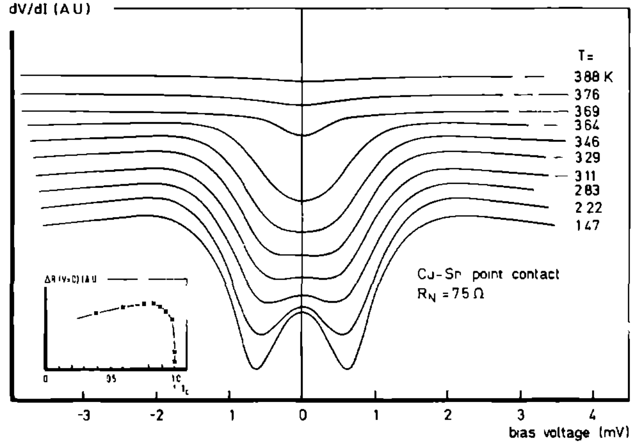


Fig. 3. dV/dI (V) characteristics for Sn-Cu point contact for varying temperatures. R_N denotes the normal state resistance measured above T_c. The vertical scale has been displaced by a constant shift for consecutive curves. The inset shows the resistance at zero voltage $\Delta R=R_N-R(V=0)$ as a function of reduced temperature T/T_c (T_c: critical temperature).

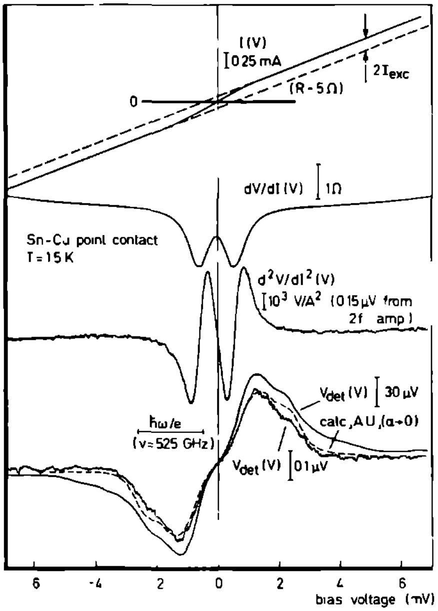


Fig. 4. Complete set of data for a representative ScN contact (Sn-Cu). Upper three curves: I(V), dV/dI (V) and d^2V/dI^2 (V) curves respectively. Lowest set curves: laser induced signal (V_{det}) as a function of bias voltage in the limit of low power and for a higher power level at a frequency of 525 GHz. A calculated result in the limit of low power is given by the dashed curve.

retical approach chosen by Klapwijk et al.[8] is more practical to analyze experimental results.

To explain the rather strong tunneling character of the high Ohmic junctions used by Gubankov et al.[9], it was assumed that in addition to a very small constriction, which was assumed to fulfil the conditions for the theory of Artemenko et al.[6] (dirty case), the surrounding oxide layer provided a parallel conducting path. The parallel connection of a tunneling conductance (oxide layer) with a shorting constriction has been used to model ScS point contacts with widely varying resistances also[19].

To explain his results on very low Ohmic ScN junctions, Dorozhkin[10] has argued that the theory of Zaitsev[7] qualitatively may have a much larger range of validity and will hold also for constrictions satisfying $\ell << a << \ell_\varepsilon$. Here ℓ_ε is the energy relaxation length $\ell_\varepsilon = (\ell_o \ell_{ph})^{\frac{1}{2}}$ with ℓ_o impurity scattering length and ℓ_{ph} the electron-phonon scattering length.

It seems that the typical contacts investigated in the present work most closely meet the conditions of Zaitsevs theory: the contacts of Gubankov et al.[9] are either dirty or of the tunneling type, while the requirement $\ell >> a$ is not met in the contacts of Dorozhkin[10]. With the additional assumption of a highly, but not perfectly, transparent barrier as in the theory of Klapwijk et al.[8] the contacts can be qualitatively described very well. (See also ref. 11.)

As a useful quantitative parameter to characterize the experimental data, might be taken the ratio $s \equiv R_d(V >> \Delta)/R_{d,min}$, where $R_{d,min}$ is the minimum value of R_d. This minimum occurs at V=0 for barrierless contacts and for $V=\pm\Delta/e$ for contacts with finite barrier. In the first case s=2 at T=0 theoretically, independent of resistance, while in the latter case it depends on Z. For very low values of Z, s depends only weakly on Z[8] and will still be close to 2. In fact, at T=0, s increases somewhat with increasing Z but this increase will be opposed by the finite temperature smearing. Thus the experimentally determined parameter s may be compared to the theoretical value of 2. The experimental value, as observed for several tens of contacts, ranges between 1.2 and 1.5 in the resistance range from ∿ 5 to ∿ 100 Ω, without an obvious dependence on resistance. Within this large scattering range, no typical differences existed whether Nb or Sn was used as the superconductor. This value is much lower than the theoretical value, but lies in the same range as previously observed by Dorozhkin[8] for Al, where it ranged from 1.2 to 1.7 in the resistance range from 0.2 - 2 Ω. (In the latter work, this value was actually compared with 2.91, which is obtained from Zaitsevs incorrect factor 3, including temperature correction.) Fig. 4 shows a complete set of data for a representative ScN (Sn-Cu) contact. The excess current is clearly present in the I(V) characteristic. The measured value $I_{exc}eR/\Delta$ equals 0.9 where the value of Δ obtained from the dips in the $R_d(V)$ curve (0.58 meV) agrees well with the bulk value for Sn at $T << T_c$. The obtained value for $I_{exc}eR/\Delta$ is larger than the theoretical value for a "dirty" contact $((\pi^2/4-1)/2 = 0.73)$[6] but considerably smaller than the maximum theoretical value for a pure, barrierless contact $(4/3 = 1.33)$[7,8]. The value 0.9 corresponds to a value $Z \sim 0.4$ in the theory of ref. 8, which is very well consistent with the magnitude of the relative resistance maximum $R_d(V=0)$. When multiplying $I_{exc}eR/\Delta$ with a factor of 2 to compare with the values of Khotkevich et al.[5] for ScS contacts, it may again be concluded independently that the contact is reasonably, but certainly not optimally, "clean".

2. 2 ii. Nonlinearities due to supercurrent quenching.

The peaks appearing in the contacts of type II are not contained in the microscopic theories[6-8]. They reflect the quenching of the superconductor so that the excess current disappears and the I(V) characteristic jumps over to an Ohmic behaviour. Although usually continuous, as in Fig. 2, it occurs sometimes as a discontinuous jump which is slightly hysteretic. Quenching of superconductivity occurs when the local critical current density I_c is exceeded. I_c may itself depend on the current because of the increase in the local temperature due to Joule heating which lowers I_c. Iwanyshyn and Smith[20] have performed detailed calculations of $I_c(T_c(\vec{r}))$. They showed that in an ScS contact a normal region will spread out to occupy a larger volume near the contact with increasing current. This effect will lead to a gradual increase in differential resistance. In fact, this model alone will also lead to an excess current and it has been proposed to explain earlier excess current data[21] in ScS junctions. A similar process may also take place in ScN junctions. The rather sudden resistance peak however is not explained by this model. It was discussed qualitatively[20] that such sudden jumps might result from cumulative chain-reaction-like processes for a multicontact junction.

A gradual and continuous decrease of the excess current in ScS contacts was also observed in the recent work of Khotkevich et al.[5]. In addition however, sudden or discontinuous and hysteretic jumps also were observed, which presumably are of similar nature as those observed in the ScN contacts of the present work. The discontinuous jumps were however attributed to a nonmonotonous temperature rise with monotonously increasing bias voltage. Discontinuous temperature changes with voltage occur through the energy dependence of the mean free path. The jumps did occur in the same voltage range where the maximum nonlinearity occurs when the contact is in the normal state. The same situation seems to apply also to ScN contacts. Fig. 5 shows a series of d^2V/dI^2 vs V curves. At 3.8 K, Sn is in the normal state and the usual phonon spectra are visible[4] although somewhat smeared out due to the finite temperature. The structure near 5 and 15 mV corresponds to phonon frequencies of Sn[5]. Phonon frequencies of Cu were not observed in any of the normal state Sn-Cu contacts investigated. Although in principle the phonon frequencies of both metals may be present in the point contact spectra of heterojunctions[4], their relative magnitudes may depend on detailed contact parameters[22]. Upon decreasing the temperature, the structure in the second derivative near 15 mV gradually changes character and at the lowest temperature shown corresponds to a rather strong peak in the resistance. This peak corresponds to the disappearance of the excess current. The excess current is build up when Sn becomes superconducting (at T ∿ 3.7 K) and is characterized by the resistance dip at low voltage (or strong peak in d^2V/dI^2 near V=0). In contacts in the intermediate temperature range (e.g. T = 2.8 K), where no distinct jump occurs, it is not clear whether the Sn is still superconducting at the voltage where the phonon peak occurs. (The excess current may have been disappeared very gradually which is not easily measured in a dV/dI or d^2V/dI^2 curve; however, the Sn should be normal if I_c does decrease with temperature.) It should be noted that the structure at the low voltage phonon peak (indicated by arrows in Fig. 5) remains visible down to the lowest temperature shown. At the lowest temperature however, it is not clear whether it results from the same physical mechanism as in the normal state, or also corresponds to a partial suppression of I_{exc}. The suppression of I_{exc} in two steps was observed indeed for ScS contacts[5] where an intermediate ScN stage may occur, but is more difficult to explain for ScN contacts.

It should be emphasized that the behaviour depicted in Fig. 5 is not rigorously followed by all contacts and the resistance peak does depend on voltage (note the extreme low value of voltage in Fig. 2, type II). Some resistance dependence of this peak also was found by Khotkevich et al.[5]. The varying behaviour may be related to the microscopic nature of the contact, notably the ratio ℓ/a. In ref. 5, a correlation was also found between the strength of the nonlinearity in (1/Ra)(dR/dV) in the normal state and the magnitude of the excess current.

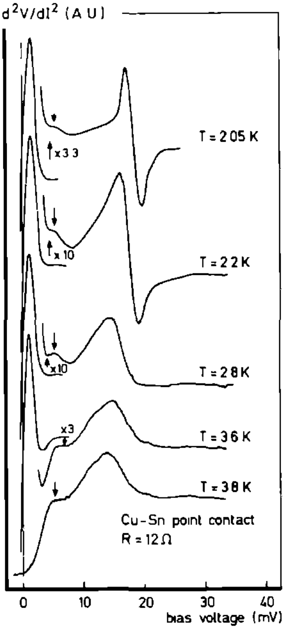


Fig. 5. Temperature evolution of d^2V/dI^2 (V) characteristic for a Sn-Cu point contact from much below T_c (upper curve) to above T_c (lower curve). Except for the lowest curve, the vertical scale is changed between low and high voltage positions. Unmarked arrows point at the position where low energy phonon structure is present in the normal state trace.

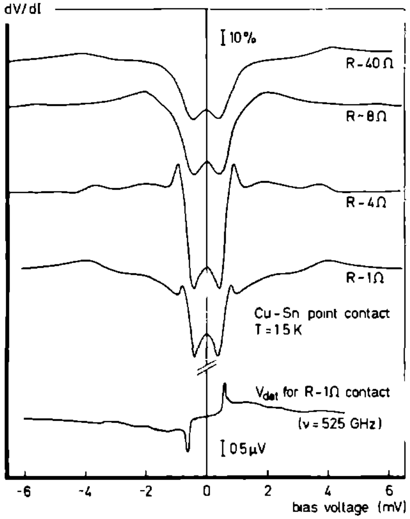


Fig. 6. Upper four curves: sequence of dV/dI (V) curves for a Sn-Cu contact obtained by increasing the contact pressure. The dV/dI-curves change character from type I to type IV from above to below (see Fig. 2). The lowest trace displays the laser induced signal (V_{det}) as a function of bias voltage obtained for the 1 Ω, type IV contact. Laser frequency: 525 GHz.

Thus as far as the current quenching effects are concerned, the ScN contact behaves very similar to the ScS contact. The invariable observation of high quality normal state spectra (see e.g. Fig. 5) as well as the continuous and well defined transition from NcN to ScN (see Fig. 5) in most cases, gives additional support to modelling the "typical" ScN contact as a clean (i.e. $\ell >> a$) orifice between the electrode metals, while a possible barrier should have high transparency.

We are now to discuss the type IV characteristic of Fig. 2. An important observation is the rather large ratio s ($\equiv R_d(V >> \Delta)/R_{d,min}$). For the actual contact, whose $R_d(V)$ curve is shown in Fig. 2, s ≃ 2 which would just be compatible with the theory, but the ratio is much larger than for contacts of type I-III. Values of s clearly larger than 2 were also frequently observed (see e.g. the contact displayed in the inset of Fig. 2). This indicates that contacts of this type are qualitatively not described by the microscopic theories.

ScN contacts having large values of s have been published previously. For instance from the I(V) characteristics between a normal metal and a type II superconductor (Nb, Ta and Nb alloys) displayed in ref. 23 for very low Ohmic junctions ($\sim 10^{-3}\ \Omega$), values of $s \sim 5$ may be inferred. Also Gubankov and Margolin[9] report "quasi-Josephson" junctions for low resistance ($R \leq 10\ \Omega$) Nb-Cu contacts. There seems to be a tendency for this type of contacts to occur predominantly for low Ohmic junctions. It was also found that Nb-Cu displayed type IV behaviour more often than Sn-Cu.

For large openings, (i.e. $\ell << a$) a simple model without theory may explain this behaviour. For $\ell << a$, the spreading resistance in the normal metal ($R_s = \rho/2a$) is much larger than the NcS contact resistance and the voltage drop exists completely in the normal metal, not across the interface. Power is dissipated in the normal metal only and also the heat can be transported mainly by the normal metal. The contact provides therefore an Ohmic resistance until the critical current of the superconductor is exceeded, driving it normal in a rather abrupt way after which the contact is again Ohmic, now having the resistance with the superconductor in the normal state. Note that there is no excess current beyond the resistance jump. A similar discussion was given in ref. 20, where the contacts were also assumed to be in the dirty limit ($R = \rho/2a$).

Although the origin for the resistance jump for type IV is similar to the jumps discussed before for types I-III, the nature of the contact is clearly different. The dependence on the ratio ℓ/a is also supported by experiments such as those displayed in Fig. 6, where the resistance of a type IV contact is gradually increased by decreasing the contact pressure. Although the resistance behaves somewhat erratic at high voltages, the 1 Ω contact is clearly of type IV. In this case however, there is a relative resistance maximum at V=0. Possibly, this maximum results from an additional tunneling resistance at the NS interface. Note that the resistance dips at both sides of the maximum are anomalously sharp and are easily distinguished from the dips of type I contacts. When the resistance increases (and presumably the ratio ℓ/a increases), the contact becomes of the normal type I. It is interesting to note that the transition from type IV to type I occurs gradually, the typical wing shape of the 8 Ω contact representing an intermediate situation. A small resistance peak near 4 mV may be seen for the 40 Ω contact so that this junction might be classified as type II rather than type I. (The lower curve in Fig. 6 will be discussed later.)

3. High-frequency effects.

3 i. Classical rectification and Photon Assisted Tunneling.

It is well known, that the point contact geometry couples very well to an applied radiation field. The wire acts hereby as an antenna to couple high-frequency (hf) signals across the contact and can be represented as a voltage source with internal impedence R_a, which is loaded by the junction. A typical value for R_a at FIR frequencies is ∿ 150 Ω so that the signal across the junction will be current controlled for junction resistances considerably less than R_a. The current through the junction will thus be modulated at the frequency of the radiation. Classically, the resulting average dc voltage across the junction will be given by:

$$V_{dc} = \frac{1}{T} \int_0^T V(I + i_o \cos \omega t)dt \qquad (5)$$

where I is the applied dc current, i_o and ω the amplitude and frequency of the hf-signal, V(I) the (dc) current-voltage characteristic, $\omega T=2\pi$. An analogous expression holds for a voltage bias. V_{dc} in eq. (5) is unequal to zero for a nonlinear V(I). For low i_o, (5) may be approximated by

$$V_{det} \equiv V_{dc} - V_o = \frac{1}{4} \frac{d^2V}{dI^2} i_o^2 \qquad (6)$$

where d^2V/dI^2 is the second derivative of V(I) at the bias point and V_o the dc-voltage without hf-signal. Eqs. (5) and (6) are the well known expressions for classical rectification. Classical rectification is independent of frequency. The hf induced voltage is thus for low signal amplitudes proportional to d^2V/dI^2, which is directly measured with audio-frequency modulation techniques (see upper three curves of Fig. 4). The lowest set of curves in Fig. 4 displays the laser induced dc-voltage at a frequency of 525 GHz. These curves are not proportional to d^2V/dI^2, which is the case down to arbitrarily low power levels.

It was previously shown[14] that the so called "Photon Assisted Tunneling" model, which may be considered as the quantummechanical analogue of classical rectification, is applicable to the data. In this model, the dc current-voltage characteristic under the influence of radiation can be written as[15,16]:

$$I_{dc}^{rad} = \sum_{n=-\infty}^{\infty} J_n^2(\alpha)\, I(V_o + n(\hbar\omega/e)) \tag{7}$$

It is assumed that the voltage across the contact is controlled and varies as $V_1 \cos\omega t$. $\alpha \equiv eV_1/\hbar\omega$ and J_n is the ordinary Bessel function of order n. V_o is the dc bias voltage. The amplitude of the dissipative part of the current at frequency ω is given by[16]

$$I_\omega(V_o) = \sum_{n=-\infty}^{\infty} J_n(\alpha)\,[\,J_{n+1}(\alpha) + J_{n-1}(\alpha)\,]\, I(V_o + n\hbar\omega/e) \tag{8}$$

For small α, the Bessel function may be expanded and (7) and (8) become

$$I_{dc}^{rad}(V) = I(V_o) - V_1{}^2\,\tfrac{1}{4}\,\frac{I(V_o+\hbar\omega/e - 2I(V_o) + I(V_o-\hbar\omega/e)}{(\hbar\omega/e)^2} \tag{7'}$$

$$I_\omega(V_o) = V_1\,\frac{I(V_o+\hbar\omega/e) - I(V_o-\hbar\omega/e)}{2\hbar\omega/e} \tag{8'}$$

When changes in I(V) are small on a voltage scale small compared to $\hbar\omega/e$, 7' and 8' can be expanded to lowest order and yield

$$I_{dc}^{rad}(V_o) = I(V_o) - \frac{1}{4}\,V_1^2\,\frac{d^2I}{dV^2} \tag{7''}$$

$$I_\omega(V_o) = \frac{dI}{dV}\,V_1 \tag{8''}$$

which are again the classical results. (7") is equivalent to (6), but now for an assumed voltage controlled contact. With the ac power dissipated in the junction given by $P = \frac{1}{2} I_\omega V_1$, the induced dc-current $\Delta I_{dc} \equiv I_{dc}^{rad} - I(V_o)$ per unit of power absorbed, S_q, becomes:

$$S_q = - \frac{e}{\hbar\omega} \left\{ \frac{I(V_o + \frac{\hbar\omega}{e}) - 2\, I(V_o) + I(V_o - \frac{\hbar\omega}{e})}{I(V_o + \frac{\hbar\omega}{e}) - I(V_o - \frac{\hbar\omega}{e})} \right\} \tag{9}$$

and the corresponding classical expression for small $\hbar\omega/e$, is:

$$S = - \frac{1}{2} \frac{d^2I}{dV^2} \Big/ \frac{dI}{dV} \tag{10}$$

Eq. (9) is illustrated in Fig. 7 for an arbitrary I(V) characteristic.

The high-frequency response is thus determined by three points on the I(V) characteristic: at V_o and $V_o \pm \hbar\omega/e$. Whenever the resistance varies considerably on a voltage scale compared to $\hbar\omega/e$, the response according to the quantummechanical model (9) may differ considerably from the classical expression (10). As can be seen from Fig. 4, the conditions for the validity

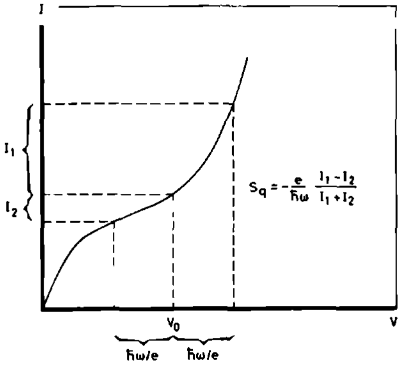


Fig. 7. Illustration of the PAT-formula (eq. (9)) for an arbitrary I(V) characteristic.

of the classical rectification model will be fulfilled only at frequencies $\lesssim$ 100 GHz for the Cu-Sn ScN point contact.

From (9) it follows that the current-sensitivity of a rectifier is always smaller than $e/\hbar\omega$, corresponding to a quantum efficiency of unity (one electron per absorbed photon). Deviations from classical rectification may therefore be expected on very general grounds, whenever the value S (eq.(10)) of an arbitrary I(V) characteristic approaches the value $e/\hbar\omega$. For typical ScN contacts studied in this work, $S \sim e/\hbar\omega$ at $\sim$ 1200 GHz.

Gubankov and Margolin[9] have recently investigated the effects of 36 GHz irradiation on the I(V) characteristics of Nb-Cu ScN contacts as a function of power. Their results were discussed in terms of the quantummechanical model described above, which, according to the theory of Artemenko et al.[6], should be valid for the specific case of the ScN contact. Based on general grounds however, no specific features of this model are expected at the low frequencies used in ref. 9 ($\hbar\omega/e \sim 0.14$ mV at 36 GHz), compared to the voltage range of the changes in $R_d(V)$ in their published characteristics. It was shown by Hamilton and Shapiro[24] that the quantum expression (7) reduces to (the voltage bias version of) the classical rectification expression (5) at arbitrary power levels. Classical rectification does contain at least qualitatively the effects observed by Gubankov et al.[9], e.g. the splitting of the $1/R_d(V)$ peak as a function of hf power level. Such effects result from "overmodulation", when the hf voltage swing V_1 is much larger than the width of the nonlinearities, and were studied previously[24,25] for superconducting tunnel junctions. It is therefore of great interest to compare the data of ref. 9 with the simple theory of classical rectification.

3 ii. Experimental results: ScN contacts.

In order to verify the quantum mechanical model in detail, experiments were performed at different frequencies, as a function of power level and for different superconductors and compared with calculations. Fig. 8 shows an $R_d(V)$ and $d^2V/dI^2(V)$ curve and the low power hf detected signal as a function of bias voltage for three different frequencies, obtained nominally on the same contact. The resistance shows a small stepwise increase at $\sim$ 15 mV after which it slowly decreases to the original value. The step corresponds to the voltage where the superconductor starts to become normal and the normal state resistance is reached near the dip at $\sim$ 28 mV. The relatively large voltage range over which the excess current is dissappearing (discon-

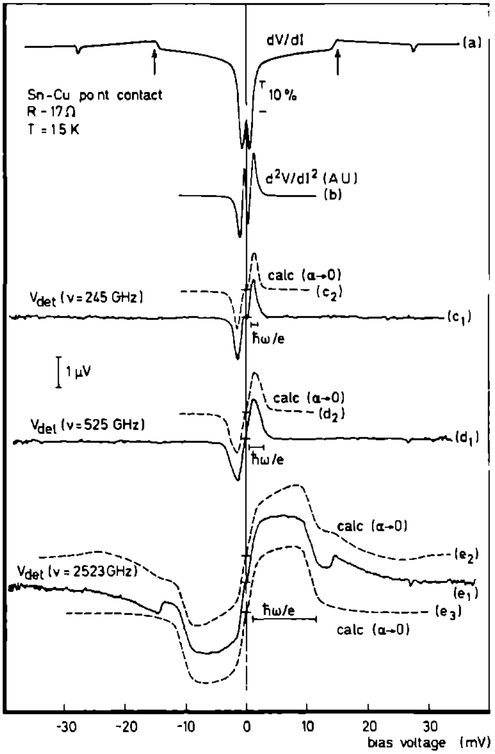


Fig. 8. (a) and (b): dV/dI (V) and d^2V/dI^2 (V) respectively as a function of bias voltage for a Sn-Cu point contact. Arrows in curve (a) point at resistance jumps due to the quenching of superconductivity. (c_1),(d_1) and (e_1): laser induced signal as a function of bias voltage in the low power limit at three different frequencies respectively.(c_2),(d_2) and (e_2): calculated responsivities in the low power limit using the dV/dI characteristic (a). (See text.) (e_3): as (e_2), but now neglecting the resistance jumps near 15 mV in the dV/dI characteristic. The data are the same as in ref. 14.

tinuously) is somewhat exceptional, but the dip, corresponding to a small "undershoot" with respect to the Ohmic I(V) curve is observed more often. Note that the resistance step occurs near the Sn phonon frequencies (compare with Fig. 5).

It is easily seen that the hf-results exhibit quantum effects, because the width of the sensitive region scales with $\hbar\omega/e$. A low resistance contact as in Fig. 8, will be current controlled with approximately constant hf current amplitudes. Under this condition, classical rectification curves would be directly proportional to d^2V/dI^2, independent of frequency. At the lowest frequency in Fig. 8 this classical behaviour is approached.

A quantum-model for a current controlled contact is not available. We have however calculated the responsivity on the basis of the voltage-controlled quantum model outlined above. This was done also in ref. 14, based only on eq. (7), with the voltage responsivity given by $\Delta I_{dc} \cdot R_d$ where $\Delta I_{dc} \equiv I_{dc}^{rad} - I(V_o)$, thereby assuming that the hf voltage amplitude remained constant with dc bias voltage. Although still within the approximation of a voltage controlled contact, a slight improvement of the model is formally expected, when power impedance mismatch variation with bias voltage (or R_d) is taken into account. The power impedance mismatch factor F_q is given by:

$$F_q = \frac{4\, R_a\, R_d^\omega}{(R_a + R_d^\omega)^2} \qquad (11)$$

with R_d^ω the (frequency dependent!) hf resistance given by V_1/I_ω (eq. 8').The power absorbed in the junction P_j is given by $P_j = F_q P$ where P is the power received by the antenna and can be assumed to be constant. The detected voltage is given by $V_{det} = -S_q P_j R_d$. For low power, R_d is practically independent of the applied power. The final expression for V_{det} thus becomes:

$$V_{det} = -S_q P_j R_d = -S_q F_q R_d \cdot P \qquad (12)$$

The $R_d(V)$ curves were stored in a computer; from these, the I(V) curves were generated (this was found to be more practical than directly storing the I(V) curves). V_{det} was calculated using (8'), (9), (11) and (12). Because the power P in (12) is unknown, the calculated curves were arbitrarily scaled to the measured curves. Because R_a is in the order of 150 Ω, $R_d^\omega \ll R_a$, so that the arbitrarily scaled results do not depend on the value of R_a.

The calculated results according to (12) are also displayed in Fig. 4 and Fig. 8, c_2, d_2 and e_2. It was found, that the use of (12) instead of the simpler expression $V_{det} = -\Delta I_{dc} R_d$ used previously, does not significantly modify the calculated results. Variations in R_d^ω , which are even smaller than variations in R_d, are small enough to be not significant. This is particularly true at the highest frequency shown in Fig. 8. These observations justify the simple approach used in ref. 14. As can be seen from Figs. 4 and 8, the calculated results describe the experimental data, which strongly differ from classical rectification, very satisfactorily. Note however that the nonlinearity of the resistance step at 15 mV is not properly treated by the quantum theory. The rather sharp step in R_d results in weak and smeared out structure in V_{det} at 2523 GHz theoretically (curve e_2, Fig. 8) and significantly influences the calculated line shape at voltages << 15 mV (15 mV - $\hbar\omega/e \sim 5$ mV) In the data however, a rather sharp structure is seen in V_{det} at the position of the step in R_d, while the line shape at low voltages ($\lesssim$ 10 mV) also differs from the calculated one.

The anomalous behaviour associated with the step structure at 15 mV reflects the macroscopic origin of the step, which results from supercurrent quenching effects. A high-frequency field does interact with this step and has the effect of reducing the voltage where it occurs. In other words: the hf-field suppresses the dc critical supercurrent . These effects were generally observed in contacts showing an abrupt step in R_d (V), also at lower frequencies. The absence of such signals near 15 mV at the frequencies 245 and 525 GHz in Fig. 8 must be accounted for by an unnoticed change in the I(V) characteristic during tuning of the laser. (Disapparance of the abrupt step in R_d(V), leaving the I(V) characteristic further practically unchanged).

Curve (e_3) in Fig. 8 is a calculated response at 2523 GHz, based upon an artificial I(V) curve which ignores the 15 mV step. For this purpose, the measured R_d(V) curve (a) below 15 mV was linearly extrapolated to the higher voltage range. Using this I(V) curve, the agreement of the calculated lineshape (e_3) with the experimental one in the low voltage regions (< 15 mV) is significantly improved. To calculate the response curve at 2523 GHz, in the region below 15 mV, the I(V) curve beyond the supercurrent quenching step is necessary in the PAT-model. It appears however that the supercurrent quenching effects as observed in a dc measurement, can be ignored at low power. (and below the dc bias voltage where it occurs).

A convenient way to identify the quantum effects also is to observe the effect of cw hf-radiation on the I(V) charactaristic (or R_d(V)) directly. Fig. 9 shows some R_d(V) curves when irradiated at different frequencies for a Cu-Sn and Cu-Nb contact. The results directly reflect the predictions of

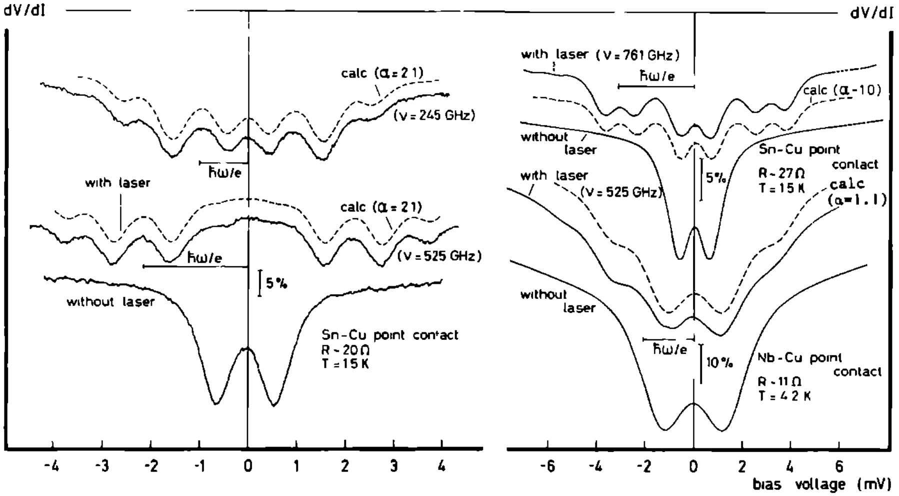


Fig. 9. dV/dI (V) characteristics for Sn-Cu and Nb-Cu point contacts with and without laser irradiation. Laser frequencies, temperatures and contact data are indicated in the Figure. Dashed curves are results calculated from the dV/dI characteristic without radiation. Values for α used in the calculations are adapted to give good agreement with the experimental curves.

eq. (7): The original $R_d(V)$ characteristic is displaced on the voltage axis by multiples of $n\hbar\omega/e$, the relative contributions weighted by $J_n^2(\alpha)$. The contributions of $n = 0$ and ± 1 are easily identified in Fig. 9, without any calculation. The calculated curves in Fig. 9 are based on eq. (7) only, using standard approximation formulas for the Bessel functions. For moderate values of α, only a few terms in the summation are significantly different from zero. (Only terms $n \leqslant 6$ in the summation needed to be retained). As discussed in connection with Fig. 8, impedance mismatch variation is not two important in the quantum model, because variations in R_d^ω are much smaller than variations in R_d. This is even more the case for the relatively high power levels employed for Fig. 9 (and R_d^ω given by (8) rather than the approximation (8')). In fact, it was observed that the I(V) characteristic at these high power levels in nearly driven Ohmic, which implies that impedance mismatch would be constant, even for classical rectification and eq. (7) only is sufficient to analyze the results. For the calculated curves, only the (constant) parameter α, which is a measure of the (a priory unknown) amount of power coupled across the contact, is adjusted to give best agreement with the data. In this way, excellent agreement with the data is obtained.

To complete the identification of the quantum effects, Fig. 10 shows some data as a function of power. The laser power was stepwise increased, but the relative attenuation was not measured. The significant feature of these curves is that the position of the laser-induced structure is determined by the frequency only, not by the power. In contrast, for classical rectification, the position of possible induced structure at high power levels does depend on power. It should be remarked that such a power independent position of the hf-induced structure is not present in the data of Gubankov et al.[9] Fig. 10(c) shows that the typical power levels used for obtaining these kind of data are sufficient to drive the contact nearly Ohmic. In the quantum-model as well as for classical rectification, the contact is driven Ohmic (in a low voltage range) at high power levels. R_d (V) curves therefore will eventually become constant, while V_{det} vs. V curves no longer exhibit quantum effects and become independent of frequency. The latter effect was clearly observed experimentally. Detection of quantum effects should therefore be done in the detection mode at low power levels (as in Fig. 8) or by investigating the dV/dI characteristic at high (but not too high) power levels (as in Figs. 9 and 10).

As was discussed in connection with Fig. 2, the so called type III contact should be considered as the most "ideal" ScN contact (that is, without

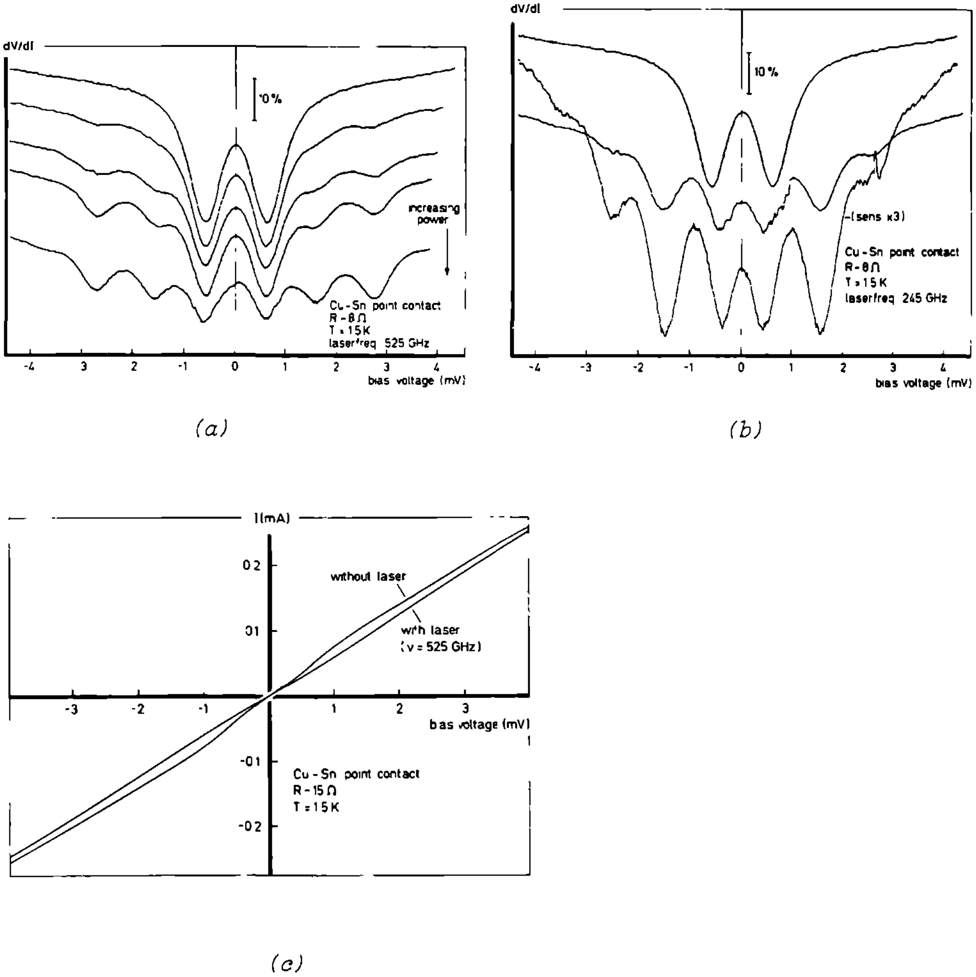


Fig. 10. (a),(b): Sequence of dV/dI-curves for a Sn-Cu point contact for increasing laser powers (uncalibrated) at the laser frequencies 525 GHz (a) and 245 GHz (b). The upper curve in (a) and (b) is the same and corresponds to the dV/dI curve without radiation. (c): Directly measured current-voltage characteristic for a Sn-Cu point contact with and without laser radiation. The power level leaving the laser is in the order of a mW.

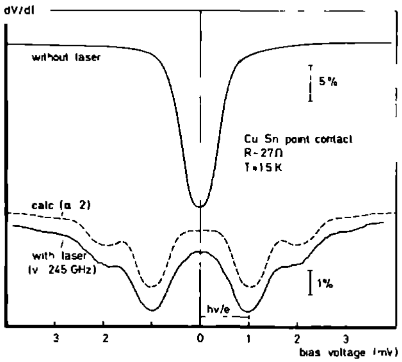


Fig. 11. dV/dI (V) characteristics for a type III contact (see Fig. 2) with and without laser irradiation. The dashed curve is calculated using the upper curve and the value of α adjusted to the data.

tunneling features). The effects of hf-interaction for such a contact is shown in Fig. 11. As expected, the agreement of the experimental data with the theoretical quantum model holds for this contact as well.

Having discussed the high-frequency behaviour of ScN contacts of type I-III, it can be seen with reference to the lower trace of Fig. 6 that the hf behaviour provides another independent way to distinguish contacts of type IV from those of type I-III. The high-frequency effects observed,are clearly not described by the quantummodel. Note,for instance,the very narrow width of the peaks in V_{det},compared to $\hbar\omega/e$ ($\sim$ 2mV) in Fig. 6. It turns out that the effect of hf-radiation is to suppress the maximum supercurrent. When the junction is dc current biased near the switching point from low resistance to high resistance (see inset Fig. 2), or near the leading edge of the $R_d(V)$ peak, the radiation switches the contact from "low" to "high" voltage behaviour with the corresponding changes in dc voltage. In fact, a similar behaviour actually is observed whereever rather abrupt switching effects are observed, e.g. $R_d(V)$ peaks of type II, whereever they occur. A weak example of this was discussed for the 2523 GHz data of Fig. 8. The suppression of the supercurrent in the superconductor near V = 0 (type IV contacts) is somewhat reminiscent to the Josephson effect. Gubankov et al.[9] have reported the observations of "quasi"-Josephson current steps (although with finite slope) for such contacts at 36 GHz. In the present experiments however, no direct evidence of the ac Josephson steps was observed, but it should be added that contacts of type I-III were selectively chosen for a more detailed hf investigation. No systematic investigation of the type IV contacts was undertaken for this work,as neither were the hf-effects on the sharp resistance peaks . The origin of the "Josephson-like" effects is clearly of interest. A real Josephson contact may occur accidentally for some contacts were two superconducting regions are connected by some weak link. A long thin superconducting channel, accidentally present, also might exhibit a real Josephson effect. For both cases, the spreading resistance of the normal metal would be in series with the Josephson contact, causing the finite slope of the current steps (including the "zero" voltage, zero power step). Such a model was suggested by Gubankov et al.[9] . This is a somewhat different situation than simply a normal region in series with a superconducting region discussed before. The lack of evidence of the ac Josephson steps, together with the similar hf behaviour for type IV contacts and type II R_d (V) peaks, suggest that the simple model is sufficient to understand the observed behaviour. Depending on the sharpness of the resistance steps or peaks, detection resulting from the

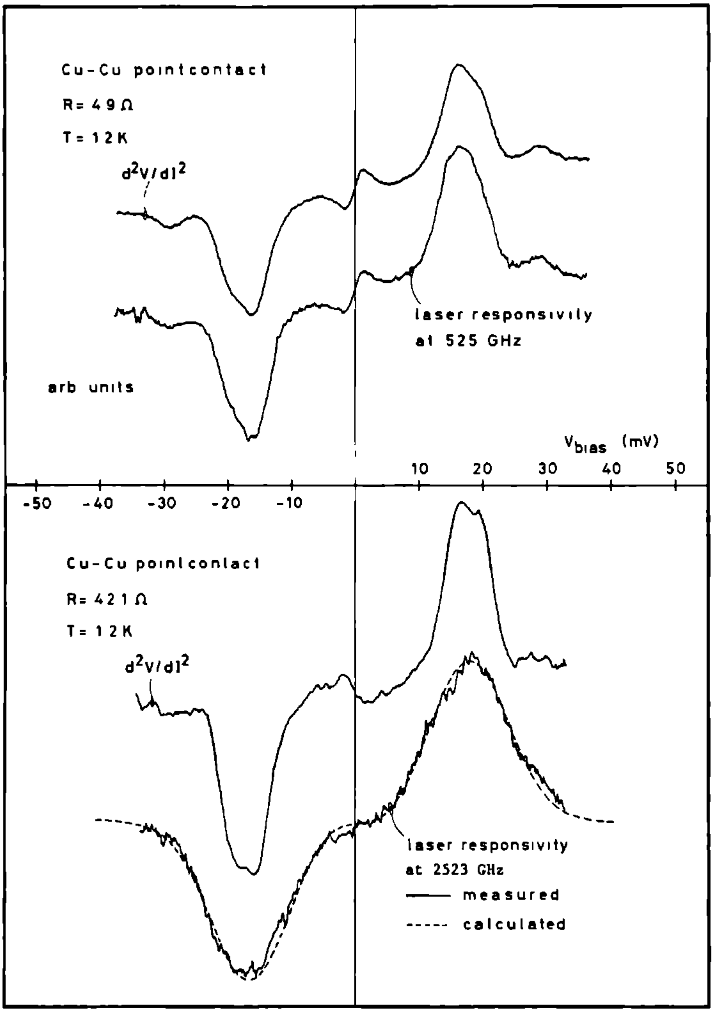


Fig. 12. Upper halve: d^2V/dI^2 (V) and laser induced signal (V_{det}) as a function of bias voltage for a Cu-Cu contact. The laser frequency is 525 GHz and is relatively low ($\hbar\omega/e \sim 2$ mV). Lower halve: as above, but for a different actual contact. The laser frequency is 2523 GHz and the value $\hbar\omega/e$ (~ 10.5 mV) is now comparable to the width of the d^2V/dI^2 signal. The dashed curve is a calculated result in the low power limit and has used the measured d^2V/dI^2 characteristic without radiation.

supercurrent suppression mechanism (either for type IV contacts or type II peaks) was found to be extremely sensitive.

3 iii. Experimental results: NcN contacts.

Thus far, the hf-interaction with ScN contacts can be explained using the measured dc I(V) characteristic only, without making explicit reference to the physical mechanism causing the particular shape of the dc I(V) characteristics. With this in mind and considering the qualitatively similar shapes of the I(V) characteristics for NcN and ScN contacts (see Fig. 1), it is of interest to examine the hf-effects for NcN contacts under conditions where the quantummodel proved to be valid for the ScN contacts. Because the nonlinear region for NcN contacts spans a voltage region at least an order of magnitude larger than for ScN contacts, quantum effects for NcN contacts will be less pronounced and detectable only at very high frequencies. Fig. 12 shows the results of hf detection for two different NcN contacts at two different frequencies, together with the measured d^2V/dI^2 curves. At the lower frequency (525 GHz) the detected signal is proportional d^2V/dI^2 and can thus be explained by classical rectification[12,13]. At 2523 GHz ($\hbar\omega/e=10.5$ meV), where the photon energy is comparable to the width of the nonlinearities on the voltage scale, a considerable broadening in the hf detected signal peaks occurs. Note that the narrow structure near the zero bias anomaly has completely disappeared. The curve at 2523 GHz was taken at sufficiently low dc output voltage, so that broadening due to hf-overmodulation[12] can be ruled out. The dashed curve in Fig. 12 is a calculated result according to $V_{det} = -\Delta I_{dc} \cdot R_d$ with ΔI_{dc} given by (7'), where the measured I(V) curve was used. As for the ScN curves, the magnitude of the calculated curve was scaled to the data. In this way, an excellent fit with the data is obtained, so that the broadening can be very well explained by the quantum model, phenomenologically. Results consistent with the quantum model also were observed at 4250 GHz, but the data were too noisy for an accurate and reliable fit. In the present experimental set-up, the antenna coupling efficiency is very poor at short wavelengths, while sensitivity of the contacts is expected to decrease at high frequencies, which makes a detailed investigation at very high frequencies difficult.

4. Discussion

The quantum theory for a nonlinear tunneling device (the so called "Photon Assisted Tunneling"-model (PAT)) was first developed by Tien and Gordon[15] for the case of single quasiparticle tunneling between superconductors. Since then the PAT effect was exclusively related to and observed in connection with the SIS (or SIN) quasiparticle tunneling devices. More recently, Tucker[16] emphasized the break down of the classical ("rectification") theory for a nonlinear element whenever the classically expected current sensitivity would approach unity quantum efficiency ($e/\hbar\omega$). He developed a full quantum theory of hf-detection and-mixing for a general single-particle tunneling device. Both the NcN and ScN contacts investigated in the present work, are characterized by the fact that both electrodes are strongly coupled, i.e. no tunneling barrier (or a very small one) is present near the contact.

From their specific theory for current transport through an ScN contact, Artemenko et al.[6] and Zaitsev[7] have derived eq. (7), for a dirty and pure contact. It is of interest to note that Artemenko et al.[6] also considered the high-frequency resistance (R_d^ω in our notation), but did not arrive at such general expression as eq. (8) (or(8')). We have assumed the validity of (8') for the ScN contact (presumably it is also implicitly in the results of Artemenko et al.) but as discussed, the results are not very sensitive to it.

Another independent derivation of eq.(7) was given by Seminozhenko et al.[26] for a specific nonlinearity in a normal metal-insulator-normal metal (NIN) tunnel junction. A nonlinearity near V = 0 ("zero bias anomaly") results here because of a nonequilibrium distribution of electrons and phonons. This nonlinearity is somewhat related to, but distinctly different from, the nonlinearities in NcN junctions discussed in the present work.

In the microscopic theory dealing with the NcN contacts, application of hf-voltages were not considered[4]. To incorporate quantum hf-effects rigorously in the theory is not straightforward, because theory is to a great extent classical. The electrons are supposed to follow classical trajectories and the kinetics are solved by Boltzmann transport equations. Under conditions $\ell > a$, with ballistic electron injection and the resistance determined by R_{Sh}, a constriction type contact is similar to a tunnel junction in many respects. The expression for R_{Sh} may be obtained from the standard expression[27] for the current through a tunnel junction, however with a tunneling probability equal to unity and a constant (i.e. energy independent) density of states[4]. As a consequence, the PAT expression (7) is valid for the field emission current, also for small constriction point contacts (governed by R_{Sh}). Although the field emission current itself depends linearly on voltage, the electron energy disequilibrium induced by it, is essential for the nonlinearity in NcN point contacts. In the microscopic theory, the nonlinearities arise from (voltage dependent) corrections to the field emission current. With the modified electron energy distribution under the action of an hf-field, as expressed by eq. (7), the corrections apply in a similar way for each term in the summation of eq. (7). The applicability of the PAT effect for the NcN contacts is therefore completely understood phenomenologically.

The PAT effect[15] is a simple case of the general frequency dependent theory for superconducting tunnel junctions due to Werthamer[28] and obtained from it by ignoring the pair current. The essential part of this theory is however the frequency dependence of the ac Josephson effect. This theory has been used previously also to analyze the experimentally observed far-infrared frequency dependence of the ac Josephson effect, investigated with constriction type point contact Josephson junctions[3]. The experimental results were in reasonable agreement with this theory, although some discrepancies were discussed in terms of additional relaxation time effects, phenomenologically accounted for by the time dependent Ginzbury-Landau theory (TDGL). Other related work on the frequency dependence of the ac Josephson effect in thin film microbridges[29] used the TDGL-theory successfully, but with somewhat unexpected parameters. The TDGL-theory has moreover been used for Josephson junctions in earlier work[30] to explain the excess current phenomenon, which is by now believed to be due to the Andreev scattering effect[6-8] and intimately connected to a constriction type contact. The general applicability of TDGL-theory is however poorly justified for ordinary superconductors with gap.

Apart from the case for superconducting quasiparticle tunnel junctions[15,28] (or general single particle tunneling[16]), the PAT effect has apparently independently been derived theoretically for some other explicit cases as well[6,7,26], including the Andreev reflection process in ScN contacts. The PAT mechanism is effective whenever field emission occurs (either in tunnelcontacts or small constriction contacts) and describes the (stationary) quantized energy distribution of the electrons in one electrode with respect to the other.

From a purely classical point of view, some high-frequency limitations of the classical picture for the NcN contact were previously given[12]. A very fundamental limitation, which presumably also underlies to the PAT-effect, is that the phase of the hf-field should not change during the passage of an electron through the field region. In constriction type contacts (NcN, ScN) the field region roughly extends over a region comparable to the orifice diameter a, so that this limit will be given by $\omega \lesssim v_F/a$. For typical values of a, this corresponds to frequencies $\lesssim 10^{14}$ Hz and is much higher than the frequencies used for the present work. For tunnel junctions, the corresponding time is the time of flight through the insulating layer which is even much shorter.

The nonlinearity of the NcN contacts is caused by electron-phonon interaction processes and is explicitly expressed[4] in terms of the energy dependence of the electron-phonon relaxation time τ_e. Clearly, intrinsic ac-effects are irrelevant as long as $\omega\tau_e < 1$[12]. Actually, τ_e is not well defined, because it depends on energy and for practical estimates, some suitable average between maximum and minimum values should be used. Making rough order of magnitude estimates, it is convenient to write[4] (see eqs. (1) and (2)):

$$\Delta R = \Delta R_s = (m^*/2ane^2) \, \Delta \, (1/\tau_e) \tag{13}$$

The energy dependence of τ_e (which at low temperatures is the electronic lifetime for spontaneous phonon emission) is the very origin of the nonlinearity. τ_e may be arbitrarily large at zero energy, while the maximum value is obtained for energies equal to the maximum phonon energies[4] after which it is constant:

$$1/\tau_{max} \sim (2\pi/3) \, \lambda_{ep} \, (k \, \Theta_D/\hbar) \tag{14}$$

with λ_{ep} the electron-phonon mass renormalization constant. Typically, $0.1 < \lambda_{ep} < 1$ for most metals, while for Cu $\lambda_{ep} \sim 0.1$. The maximum value for τ_{max}^{-1} is thus of the order $\sim k\Theta_D/\hbar$ and given roughly by the position on the voltage scale of the maximum nonlinearity of the I(V) characteristic. Upon defining an average value of $1/\tau_e$ by

$$1/\tau_e^{av} \equiv e \, \Delta V/\hbar \tag{15}$$

where ΔV is the width in the I(V) characteristic over which the changes in R occur ($\sim$ the width of the second derivative peak, $e\Delta V \leq k\Theta_D$), it is seen that the condition $\omega\tau_e^{av} < 1$ is equivalent to $\hbar\omega/e < \Delta V$, which is the condition for classical rectification as formulated from the quantum theory point of view.

Evidently, such formal redefinition of the structure width in the I(V) characteristics in terms of some time scale can be carried through also for other devices. For instance it is well known that the width of the gap-structure in SIS-or SIN-(and presumably also ScN-) junctions is ultimately limited by the quasiparticle lifetimes. To arrive at the PAT model, the assumption $\omega\tau_e > 1$ with τ_e the electron-phonon relaxation time was explicitly made in ref. 26, because it here showed up as the natural parameter. It is interesting to note that this seems to be a superfluous assumption because formally the PAT model evidently applies for $\omega\tau_e < 1$. Under the latter condition intrinsic hf-effects are not expected and the system behaves classically. The PAT formulas however also reduce to classical rectification .

The PAT effect has recently become of great interest in view of the use of SIS or SIN junctions (including the super Schottky) as very low noise detectors and mixers at millimeter wave frequencies[31]. Because of the very strong nonlinearities, attainable with these devices, performance is strongly influenced by the PAT-effect already at microwave frequencies. The PAT-effect not only leads to quantum limited sensitivity, but also to qualitatively new effects such as signal conversion gain in a heterodyne mixer mode of operation, not possible for a classical nonlinear element[32]. In this respect, it is interesting to compare some practical ultimate figures of merit for the ScN and NcN contacts with the SIS or SIN contacts. The ScN I(V) curve may roughly be approximated to consist of two straight-line sections, one with resistance $R_d = R_1$ up to $|V| = \Delta/e$ and one with $R_d = R_2$ at $|V| > \Delta/e$. For ideal ScN contacts $R_2/R_1 = 2$ according to the theory. For a dc-bias near $V_o = \Delta/e$ (the optimum bias voltage) the current sensitivity is then given by (9) as

$$S_{q(S_cN)}\Big|_{V\sim\Delta/e} = \frac{e}{\hbar\omega} \cdot \frac{R_2 - R_1}{R_2 + R_1} = \frac{e}{\hbar\omega} \cdot \frac{1}{3} \qquad (16)$$

For SIS or SIN contacts, the quantum limited value $S_q = e/\hbar\omega$ is closely approached,but for the (clean) ScN contact, the sensitivity is principally a factor of three below the value $e/\hbar\omega$. The presence of a barrier (making the ScN contact more tunneling - like) will only improve the sensitivity. Because the I(V) curve will generally not be so sharp near $V = \Delta/e$, the limiting value will be obtained at frequencies which must relatively be somewhat higher as compared to SIS (SIN) contacts. The (intrinsic) maximum frequency for quantum limited sensitivity is set by the voltage difference between the position of the maximum of the nonlinearities on both sides of zero bias. At higher frequencies, the ratio $(I_1-I_2)/(I_1+I_2)$ in (9) becomes small. Thus the frequency limits are given by $\hbar\omega \leqslant 2\Delta$ for ScN and SIN and $\hbar\omega \leqslant 4\Delta$ for SIS. (Pair breaking at $\hbar\omega \geqslant 2\Delta$ might also limit the performance of superconducting devices, but is presumably of secondary importance). Note that this limit is exceeded by an order of magnitude in the present work for Sn - Cu contacts ($2\ \Delta_{Sn} \sim 1.2$ meV) at 2523 GHz. ($\hbar\omega = 10.5$ meV). Also Josepson effect devices are frequency-limited by the superconducting energy gap value in Werthamers theory[28,33,34], but may be operated considerably above this roll-off frequency[3].

Ultimate values for the sensitivity of NcN contacts in the quantum limit are much lower and a rough estimate may be obtained from (9) using typically observed $R_2/R_1 \sim 1$-10% values, which gives

$$S_{q(NcN)} \Big|_{V \sim k\Theta_D/e} \lesssim \frac{e}{\hbar\omega} \cdot 0.1 \qquad (17)$$

(Theoretical values for R_2/R_1 are not explicit in the theory[4]). This quantum limited sensitivity will be reached at frequencies of the order of $\omega \sim k\Theta_D/\hbar$, which is an order of magnitude higher than the frequency range where the superconducting devices (including the SIS or SIN junctions) have their optimum performance.

Formally, the quantum model may be considered as providing a very fundamental and intrinsic high-frequency limitation to any nonlinear element[16]. At high enough frequencies,it produces a sensitivity roll-off at least as fast as $1/\omega$ according to (9). For nonlinear elements, where the nonlinearity is restricted to a finite portion of the I(V) curve (the ScS, NcN, SIN, SIS diodes are reasonable examples of such situation) the roll-off will become even much faster when the ratio $(I_1-I_2)/(I_1+I_2)$ in (9) becomes negligibly small.

The highest frequency operation with a nonlinear device has been obtained in metal-insulator-metal tunnel junctions operating at room temperature (conventionally designated as "MIM"). They are either constructed in the form of point contacts or as evaporated thin film devices. Although the nonlinearity of these devices is usually very small, values $S \leq 10\ V^{-1}$ can be obtained. Greatest interest in these devices is at infrared or even visible light frequencies ($\hbar\omega \sim 2$eV). Classical rectification is well accepted as the operation mechanism of these devices up to infrared frequencies (~ 30 THz). Although classical rectification has been observed even at visible light frequencies[35], it is mostly found that classical rectification does not apply at such high frequencies[36,37]. It is interesting to note,that in some cases the conditions for the quantum effects are, in fact, fulfilled[16] ($S \geq e/\hbar\omega$, see e.g. refs. 36, 37, 38).

From the experimental data, it was concluded that internal photo-emission was responsible for visible light detection in an evaporated tunnel junction[37]. In such a process, an electron in one of the electrodes is photo-excited to an energy $\hbar\omega$ above its initial energy and the photon energy is larger than the maximum barrier height ϕ_{max} of the insulator. Electrons with energy $E > \phi_{max}$ can drift from one electrode to the other via the conduction band of the insulator. This is an example of a photoconductivity process. Photo-excitation will also lead to a photocurrent, even when $\hbar\omega < \phi_{max}$. Because the tunneling probability in general is energy dependent, excited carriers easier tunnel through the barrier than equilibrium carriers. The energy dependence of the tunneling probability is significant however only at the energy scale of ϕ_{max} (1-2 eV typical), thus at high (near-visible) frequencies. This process is known as Photo-Induced Tunnel Currents (PITC)[39]. Sometimes it is also quite naturally referred to as photon assisted tunneling[37].

There are some very striking correspondences between the PITC-(or even internal photo-emission) effect and the PAT-effect discussed in this work, but yet they seem to be distinctly different in nature. At low power levels, the net effect in both cases can be described as an increase in population of states in the energy range $\hbar\omega$ above the Fermi level, at the cost of an equal decrease in population in the energy range $E_F-\hbar\omega$. (It is interesting to note that the same illustrations were used to explain the PITC-effect in ref. 39 and the PAT-effect in ref. 26.) For the PITC-effect, an electron is first photo-excited in one electrode near the barrier, and subsequently tunnels across the barrier. For this purpose, at least one of the thin film electrodes is made thin enough to be transparent[39] or the light is focussed at a thin edge of the junction[37]. In the PAT-effect, the (hf-induced) dc current results from the same field emission process responsible for the dc current, but at high-frequency ac voltages, the voltage oscillations should be considered as "quantized" according to the PAT-theory. A higher level of theoretical understanding, which should provide a more unified description of the PITC- and PAT-effect would be desired

5. Conclusion.

The dc and ac behaviour of two types of constriction type point contacts in the clean limit, $\ell \gg a$ (ScN and NcN), were investigated. The nonlinearity in ScN contacts is dominated by the excess current phenomenon, which was recently explained in terms of Andreev reflection processes[6-8]. The contacts studied, have a character intermediate between ScN contacts of previous investigations[9,10] (see also ref. 11). The detailed shape of the I(V) characteristic near V=0 varies from contact to contact in an unpredictable way. This is qualitatively consistent with the theory of Klapwijk et al.[8], where it was found that the shape of the I(V) characteristic sensitively depends on the magnitude of the barrier potential used to characterize the contact. This barrier however always has a transparency (>> 0.8) high enough that it does not seriously affect the behaviour of NcN contacts. The correlation found by Khotkevich et al.[5], with the supercurrent quenching effects in ScS contacts and their behaviour in the normal state, is partly corroborated by the present experiments with ScN contacts.

The far-infrared response of the ScN contacts can be described using the observed dc I(V) characteristic only, without explicit reference to the physical mechanism behind it. The results are in good agreement with the model of "Photon Assisted Tunneling" as expected from the theory[6,7].

Strong indication exists that the same model also applies to the high-frequency behaviour of NcN contacts. This implies the first observation of the PAT-effect in a nonlinear element where no superconductor is involved.

Acknowledgements.

We are most grateful to Dr. A.P. van Gelder, Dr. A.G.M. Jansen and Dr. H. van Kempen for clarifying discussions and to Dr. T.M. Klapwijk for preprints of ref. 8 prior to publication and interesting discussions. We also thank G.E. Blonder and M. Tinkham for a preprint of ref. 11 prior to publication.

REFERENCES

1. D.G. McDonald, V.E. Kose, K.M. Evenson, J.S. Wells and J.D. Cupp, Appl. Phys. Lett. 15, 121 (1969); D.G. McDonald, K.M. Evenson, J.S. Wells and J.D. Cupp, J. Appl. Phys. 42, 179 (1971)
2. J.E. Zimmerman, Proc. Appl. Superconductivity Conf., Annapolis, 1970, p. 544 (unpublished)
3. D.A. Weitz, W.J. Skocpol and M. Tinkham, Phys. Rev. Lett. 40, 253 (1978); D.A. Weitz, W.J. Skocpol and M. Tinkham, Phys. Rev. B 18, 3282 (1978)
4. For a review, see A.G.M. Jansen, A.P. van Gelder and P. Wyder, J. Phys. C: Solid State Phys. 13, 6073 (1980)
5. A.V. Khotkevich and I.K. Yanson, Fiz. Nizk. Temp. 7, 727 (1981) [Sov. J. Low Temp. Phys. 7, 354 (1981)]
6. S.N. Artemenko, A.F. Volkov and A.V. Zaitsev, Solid State Commun, 30, 771 (1979)
7. A.V. Zaitsev, Zh. Eksp. Teor. Fiz. 78, 221 (1980) [Sov. Phys. JETP 51, 111 (1980)]
8. T.M. Klapwijk, G.E. Blonder and M. Tinkham, Physica 109B (1982) (preprint); G.E. Blonder, M. Tinkham and T.M. Klapwijk, Phys. Rev. B 25, 4515 (1982)
9. V.N. Gubankov and N.M. Margolin, Pis'ma Zh. Eksp. Teor. Fiz. 29, 733 (1979) [JETP Lett. 29, 673 (1979)]; V.N. Gubankov and N.M. Margolin, Zh. Eksp. Teor. Fiz. 80, 1419 (1981) [Sov. Phys. JETP 53, 727 (1981)]
10. S.I. Dorozhkin, Zh. Eksp. Teor. Fiz. 79, 1025 (1980) [Sov. Phys. JETP 52, 520 (1980)]
11. G.E. Blonder and M. Tinkham, preprint, 1982
12. R.W. van der Heijden, A.G.M. Jansen, J.H.M. Stoelinga, H.M. Swartjes and P. Wyder, Appl. Phys. Lett. 37, 245 (1980)
13. R.W. van der Heijden, H.M. Swartjes and P. Wyder, unpublished
14. R.W. van der Heijden, J.H.M. Stoelinga, H.M. Swartjes and P. Wyder, Solid State Commun. 39, 133 (1981)
15. P.K. Tien and J.P. Gordon, Phys. Rev. 129, 647 (1963)
16. J.R. Tucker and M.F. Millea, Appl. Phys. Lett. 33, 611 (1978); J.R. Tucker, IEEE J. Quantum Electron. QE-15, 1234 (1979)
17. Actually a factor of 3 is found in ref. 7. It can be seen from eq. (45) of ref. 7 that the discrepancy results from the same term which leads to the expression for I_{exc}, which seems to be incorrect by a factor of 2; see remarks in ref. 8 and footnote in ref. 5. Correcting the excess

current term in ref. 7 by a factor 2 also leads then to a resistance ratio of 2.

18. M.K. Chien and D.E. Farrel, J. Low Temp. Phys. 19, 75 (1975)
19. Yu.Ya. Divin and F.Ya. Nad', Fiz. Nizk. Temp. 4, 1105 (1978) [Sov. J. Low Temp. Phys. 4, 520 (1978)]
20. O. Iwanyshyn and H.J.T. Smith, Phys. Rev. B 6, 120 (1972)
21. J.I. Pankove, Phys. Lett. 21, 406 (1966)
22. A.G. Batrak and I.K. Yanson, Fiz. Nizk. Temp. 5, 1404 (1979) [Sov. J. Low Temp. Phys. 5, 663 (1979)]
23. D.B. Sullivan and C.E. Roos, Phys. Rev. Lett. 18, 212 (1967)
24. C.A. Hamilton and S. Shapiro, Phys. Rev. B 2, 4494 (1970)
25. V.A. Tulin, Fiz. Nizk. Temp. 2, 1522 (1976) [Sov. J. Low Temp. Phys. 2, 741 (1976)]; E.D. Dahlberg, R.L. Orbach and I. Schuller, J. Low Temp. Phys, 36, 367 (1979)
26. V.P. Seminozhenko and A.A. Yatsenko, Fiz. Nizk. Temp. 5, 1118 (1979) [Sov. J. Low Temp. Phys. 5, 527 (1979)]
27. C.B. Duke, *Tunneling in Solids* (Academic, New York, 1969)
28. N.R. Werthamer, Phys. Rev. 147, 255 (1966)
29. T.D. Clark and P.E. Lindelof, Phys. Rev. Lett. 37, 368 (1976)
30. K.K. Likharev and L.A. Yakobson, Zh. Eksp. Teor. Fiz. 68, 1150 (1975) [Sov. Phys. JETP 41, 570 (1976)]
31. For a review see: P.L. Richards and T.-M. Shen, IEEE Trans. Electron Devices ED-27, 1909 (1980)
32. W.R. McGrath, P.L. Richards, A.D. Smith, H. van Kempen, R.A. Batchelor, D.E. Prober and P. Santhanam, Appl. Phys. Lett. 39, 655 (1981)
33. P.L. Richards in *Semiconductors and Semimetals*, Vol. 12, R.K. Willardson and A.C. Beer, Eds. (Academic Press, New York, 1977) pp. 395-439
34. D.G. McDonald, R.L. Petersen, C.A. Hamilton, R.E. Harris and R.L. Kautz, IEEE Trans. Electron Devices ED-27, 1945 (1980)
35. S.M. Faris, T.K. Gustafson and J.C. Wiesner, IEEE J. Quantum Electron. QE-9, 737 (1973)
36. B.L. Twu and S.E. Schwarz, Appl. Phys. Lett. 25, 595 (1974)
37. G.M. Elchinger, A. Sanchez, C.F. Davis, Jr. and A. Javan, J. Appl. Phys. 47, 591 (1976)
38. T.K. Gustafson, R.V. Schmidt and J.R. Perucca, Appl. Phys. Lett. 24, 620 (1974)
39. Z. Burshtein and J. Levinson, Phys. Rev. B 12, 3453 (1975)

SUMMARY.

In this thesis, an experimental investigation is described of the high-frequency conductivity at far-infrared frequencies of (semi-)metals under condition where quantummechanical effects are important. Two different physical systems have been chosen as probe system for this purpose, while the experimental methods in both cases differ considerably as well. In the first case, absorption measurements have been made for bulk samples as a function of external magnetic field. A point contact configuration was used for the other case, where a signal due to rectification of the far-infrared frequency was directly measured as a function of an applied electric field (voltage).

In Chapter II, a survey is given of the experimental methods used and the experimental set-up is described.

Chapter III deals with the far-infrared absorption measurements of the semimetals Bi and Sb in high magnetic fields. It was investigated how the by itself well known magnetoplasma effects (cyclotron resonance, hybrid resonance, dielectric anomalies, etc.), which can be described with simple, classical theories, evolve under conditions where the electronic system is strongly quantized. The so called quantum limit, where only the lowest Landau level is left occupied below the Fermi level, can be attained in Bi in relatively modest magnetic fields. The measurements were compared with calculations, where the bandstructure was sufficiently detailed taken into account (in the so called two band model). The magnetoplasma effects in the quantum limit turn out to be well described by elementary expressions from quantummechanical perturbation theory.

Chapter IV treats the mechanism of rectification of high-frequency voltages which are generated across a point contact by an external radiation field. In particular, point contacts between normal metals at low temperatures of the constriction type, have been used for this purpose. The I(V) characteristic of such a contact is nonlinear due to an electron-phonon scattering mechanism.

Rectification classically is independent of the frequency and depends only on the curvature of the I(V) characteristic (d^2V/dI^2). At high frequencies and/or strong nonlinearities, deviations from the classical behaviour are expected, whenever the (current-)sensitivity approaches the quantum efficiency of one electron per photon ($e/\hbar\omega$). This quantum behaviour is investigated in Chapter V for normal metal-superconductor point contacts, where the nonlinearities are caused by the Andreev-reflection mechanism. The anomalous

behaviour observed for point contacts between normal metals at the highest far-infrared frequencies is assigned to the same quantum effects as well.

SAMENVATTING.

Dit proefschrift beschrijft een experimenteel onderzoek naar de geleidingseigenschappen bij ver-infrarood frekwenties van (half-)metalen onder kondities waarbij kwantummechanische effekten een belangrijke rol spelen. Twee verschillende fysische systemen zijn voor dit doel als proefsysteem gekozen, terwijl de gebruikte experimentele methode in beide gevallen eveneens zeer verschilt. In het eerste geval werden absorptie metingen verricht aan bulk samples als funktie van een uitwendig magneetveld. Een puntkontakt konfiguratie werd gebruikt voor het andere geval, waarbij direkt een signaal ten gevolge van het gelijkrichten van de ver-infrarood frekwentie werd gemeten als funktie van een aangelegd elektrisch veld (spanning).

In Hoofdstuk II wordt een overzicht gegeven van de gebruikte experimentele methoden en wordt de opstelling beschreven.

Hoofdstuk III behandelt de ver-infrarood absorptie metingen aan de halfmetalen Bi en Sb in hoge magneetvelden. Onderzocht werd hoe de op zich bekende magnetoplasma effekten (cyclotron resonantie, hybride resonantie, diëlektrische anomalieën, etc.), die met eenvoudige klassieke theorieën te beschrijven zijn, zich gedragen onder kondities waarbij het elektronen systeem sterk gekwantiseerd is. De zogenaamde kwantum limiet, waarbij alleen het laagste Landau nivo zich beneden het Fermi nivo bevindt, is in Bi te bereiken in relatief lage velden. De metingen werden vergeleken met berekeningen waarbij de bandstruktuur redelijk gedetailleerd werd meegenomen (in het zgn. twee banden model). De magnetoplasma effekten in de kwantum limiet blijken goed te beschrijven met elementaire theorieën uit de kwantummechanische storingsrekening.

Hoofdstuk IV gaat in op het mechanisme van gelijkrichten van hoogfrekwente spanningen die door een extern stralingsveld worden gegenereerd over een puntkontakt. In het bijzonder worden hiervoor puntkontakten gebruikt tussen normale metalen bij lage temperaturen, die van het konstriktie-type zijn. De I(V) karakteristiek van een dergelijk kontakt is niet-lineair ten gevolge van een elektron-fonon verstrooiingsmechanisme.

Gelijkrichting is klassiek onafhankelijk van de frekwentie en alleen afhankelijk van de kromming van de I(V) karakteristiek (d^2V/dI^2). Bij hoge frekwenties en/of sterke niet-lineariteiten zijn afwijkingen van het klassieke gedrag te verwachten wanneer de (stroom-)gevoeligheid de kwantum efficiëncy van één elektron per foton ($e/\hbar\omega$) nadert. Dit kwantum gedrag is onderzocht in Hoofdstuk V voor normaal metaal-supergeleider puntkontakten,

waar de niet-lineariteit veroorzaakt wordt door het Andreev-reflektie mechanisme. Het afwijkende gedrag van puntkontakten tussen normale metalen bij de hoogste ver-infrarood frekwenties wordt eveneens toegeschreven aan dezelfde kwantum effekten.

CURRICULUM VITAE

R.W. van der Heijden

Geboren 5 januari 1953 te Oss.

1965 - 1971 : Gymnasium β, Titus Brandsmalyceum, Oss.

1971 : Begin studie natuurkunde, K.U. Nijmegen.

14 maart 1974 : Kandidaatsexamen natuurkunde (N2), K.U. Nijmegen.

12 mei 1977 Doctoraalexamen experimentele natuurkunde, K.U. Nijmegen (cum laude). Hoofdvak Vaste Stoffysika.

1977 - 1982 In dienst van de Stichting voor Fundamenteel Onderzoek der Materie (FOM). Promotieonderzoek op de afdeling Experimentele Natuurkunde 4, Fakulteit der Wiskunde en Natuurwetenschappen, Katholieke Universiteit Nijmegen, onder leiding van Prof.Dr. P. Wyder.

STELLINGEN

I

Voor het verstrooiingsproces van geleidingselectronen in gedeformeerde metalen moeten lange rijen vacatures, die ontstaan wanneer klimmende dislocaties elkaar snijden, tot de anisotrope verstrooiingskernen gerekend worden.

J. Friedel, Dislocations (Pergamon Press, Oxford, 1964); R.J.M. van Vucht, G.F.A. van de Walle, H. van Kempen en P. Wyder, to be published.

II

Het is mogelijk om de eventuele aanwezigheid van fractionele-lading onzuiverheden in ultrazuivere halfgeleiders te onderzoeken met behulp van fotothermische ionisatiespectroscopie.

M.J.H. van de Steeg, Proefschrift, Nijmegen, 1982.

III

Het gebruik door Bluyssen en Van Ruyven van de van Dingle afkomstige uitdrukking voor de magneetveldafhankelijkheid van de toestandsdichtheid in een metaal voor eindige relaxatietijden bij de analyse van hun metingen van de golflengteverschuiving van een DH injectielaser in een magneetveld is twijfelachtig. Dit zou mogelijk de discrepantie kunnen veroorzaken tussen de berekende afhankelijkheid van de ladingsdragerconcentratie bij de drempel en het gedrag van de drempelstroom als funktie van het magneetveld.

H.J.A. Bluyssen en L.J. van Ruyven, IEEE J. Quantum Electron. QE-17, 2190 (1981); R.B. Dingle, Proc. Roy. Soc. (London) A211, 517 (1952).

IV

De weerstands dc-SQUID kan direct gebruikt worden in heterodyne toepassingen om ver-infrarood frequenties te converteren naar geschikte lage frequenties, zonder gebruik van externe locale oscillator.

J.M.V. Verschueren, Scriptie, Nijmegen, 1982; B.T. Ulrich, R.W. van der Heijden en J.M.V. Verschueren, J. Phys. (Paris) 39, C6-1221 (1978).

V

Door gebruik te maken van de bijzondere eigenschappen van sommige chiraal smectische vloeibare kristallen, kan men een optisch tralie maken waarvan de roosterconstante met behulp van temperatuur en magneetveld continu gevariëerd kan worden.

I. Muševič, B. Žekš, R. Blinc, Th. Rasing en P. Wyder, Phys. Rev. Lett. 48, 192 (1982).

VI

Puntkontakten tussen metalen met magnetische verontreinigingen bij concentraties in het spin-glas gebied bieden de mogelijkheid de interne velden in deze systemen experimenteel te bepalen.

N. d'Ambrumenil en R.M. White, J. Appl. Phys. 53, *2052 (1982).*

VII

In Bi the electrons almost fill a Brillouin zone, there being a small number of electrons overlapping into the next zone and an equal number of positive holes. This number we estimate to be about 10^{-4} per atom.

N.F. Mott and H. Jones, The Theory of the Properties of Metals and Alloys (Clarendon Press, Oxford, England, 1936).

VIII

De poging door Gubankov en Margolin om de toepasbaarheid van het Photon Assisted Tunneling model voor normaal metaal-supergeleider puntkontakten te verifiëren bij frequenties van 36 GHz is niet gerechtvaardigd.

V.N. Gubankov en N.M. Margolin, Sov. Phys. - JETP 53, *727 (1981).*

IX

Ter vergroting van de politieke stabiliteit in Nederland dient de frequentie van verkiezingen laag gehouden te worden, ondermeer door verkiezingen voor landelijke en regionale vertegenwoordigingslichamen zoveel mogelijk op dezelfde dag te doen plaatsvinden.

Nijmegen, 8 oktober 1982 R.W. van der Heijden.